Autonomy IDOL

Mathematics, marketing, and magic make Autonomy the leader in enterprise search and content processing.

What is Autonomy? It is not an enterprise search system. Search is one component of collection of software that is middleware. Autonomy put a "black box" at the core of its software and proceeded to point out that advanced mathematics could be used to solve thorny business problems. With the advanced methods in the "black box," Autonomy operated automatically, unlike any other enterprise software. The Autonomy value proposition was that the "black box" could understand the meaning of data and information. Thus, the automatic discovery and delivery of unstructured content, and the integration of that content into applications and other enterprise systems delivered significant value.

Did Autonomy crack the code for search and content processing? Is the Autonomy approach the high-water mark in search and content processing? Does Autonomy's system work equally well on small collections of content as well on vast flows of data? Does Autonomy's system apply to fraud detection and eDiscovery as well as basic keyword search?

Autonomy was purchased by Hewlett Packard for $11 billion in 2011. A year later, HP accused Autonomy of misrepresenting its technology, revenue, and system. Autonomy was the leader in the relatively small world of enterprise search in terms of revenues, numbers of customers, and disruption of what has been a slow-moving, somewhat lackluster market sector. One thing is certain: No other vendor has been able to duplicate Autonomy's revenue success.

Math, pmanagement, marketing and magic were key ingredients in Autonomy's success. Under Dr. Michael Lynch's guidance, Autonomy was a success. Does the system work? You can test the system and decide for yourself. If you resent Autonomy's business acumen, ignore the company. Case closed.

Author's note:

This is a late 2007 draft. It will not be updated.

Stephen E Arnold, January 23, 2014

Introduction

Quite a few consultants and analysts have criticized Autonomy plc; for example, Deutsche Bank said, Autonomy's sales team are hunters, not farmers."[1]. One "real" consulting firm launched a "The Mouse That Roared" assault on Autonomy early in the company's history.

A London-based investment house funded an analysis in late 2007 focused on the company's business practices.

Table 1: Autonomy IDOL: A Bird's Eye View

Product Thumbnail	
1 Search Brand	IDOL, Meridio, Virage, Liquid Paper, Interwoven, Verity K2, Ultraseek
2 OS Supported	HP-UX, Solaris, Linux, Microsoft Windows, and any other POSIX compliant UNIX
3 Est License Fee	Typical license fee is $1 million. A custom price quote is required.
4 Functions	Content processing via automated integrated data operating layer or IDOL
5 Claimed Features	Virtually every search, content processing, rich media management, archiving, fraud detection, and electronic discovery feature is available from Autonomy
6 Downsides	IDOL is a black box. Components are a "collection of parts." Complex, expensive to maintain, difficult to manage
7 Similar To	Endeca, OpenText LiveLink, Oracle SES and middleware
Product Close Up 	Autonomy Corporation plc is a global leader in infrastructure software for the enterprise and is spearheading the meaning-based computing movement. Autonomy's technology forms a conceptual and contextual understanding of any piece of electronic data including unstructured information, be it text, email, voice or video. Autonomy's software powers the full spectrum of mission-critical enterprise applications including information access technology, BI, CRM, KM, call center solutions, rich media management, compliance and litigation solutions and security applications, and is recognized by industry analysts as the clear leader in enterprise search.

In 2002, Autonomy recapitalized. Autonomy had shifted from listing on a US stock exchange to a European exchange. Autonomy adjusted its financials. Results were stated in dollars and then in British pounds. Autonomy changed accounting firms from Ernst & Young to Deloitte Touche in 2005. Autonomy kept a close eye on the presentations that included criticism of its products, its sales methods, and its senior management. Autonomy took steps to make sure its publicity was distributed widely and IDOL included in consultants' reports. Autonomy, like Google, saw an opportunity and drove forward. By late 2007, like it or not, Autonomy is the largest enterprise content processing vendor. Autonomy's fees are among the highest in the content processing sector. Autonomy dominates its targeted markets.

1. David Woodward, "Mike Lynch," Director at http://www.director.co.uk/MAGAZINE/2011/5_May/mike-lynch_64_09.html

"My background is not particularly well off, and because of that I didn't really know a lot about professions you didn't come across every day. I didn't know, quite frankly, what an accountant did. I wanted to be a research scientist, because I had read about those, so I went up to Cambridge and did sciences."—Dr. Michael Lynch. Source: Network IT Week, October 9, 1998.

Autonomy's approach attracts much criticism, but when compared to also rans like Convera, Delphes, and Entopia, Autonomy knows how to grow. In 2003, Autonomy began referring to itself as the "de facto standard for government and intelligence processing." Competitors continue to lack Autonomy's boldness, aggressiveness, and confidence in the face of criticism.

The motives behind these negative evaluations are a follow on from the controversies Bayesian methods have triggered for centuries.

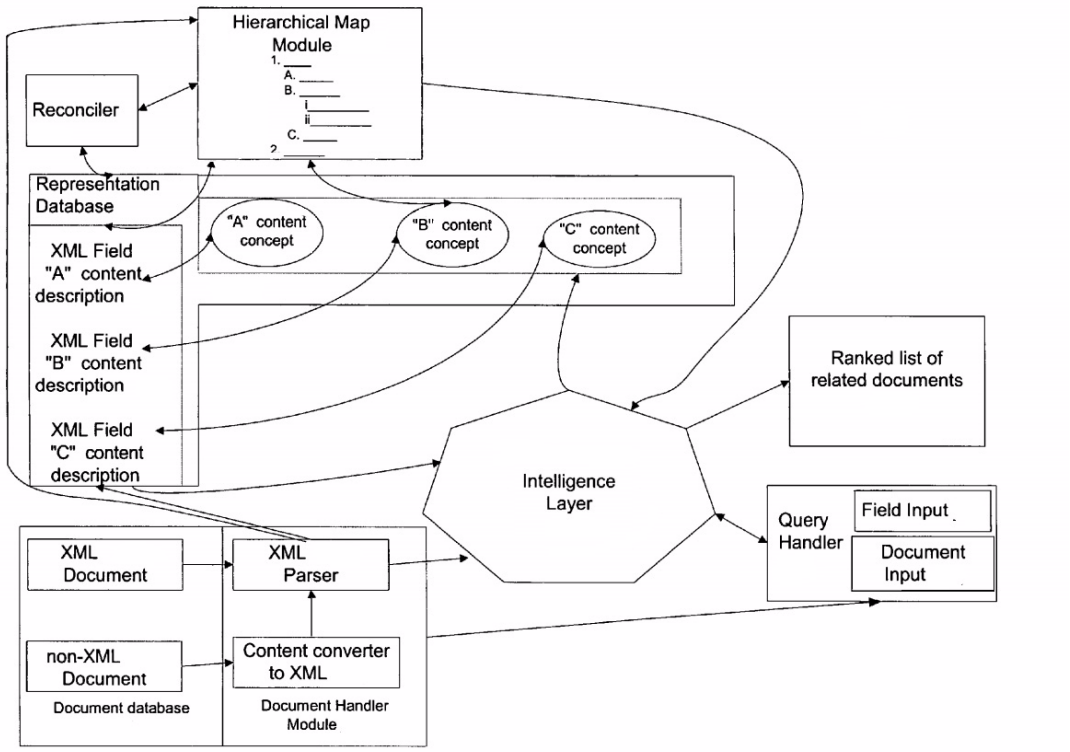
Autonomy's management seized upon the "automatic" function of Bayesian and other numerical recipes. Competitors' search and content processing systems were not positioned as "automatic." From 1996 when Autonomy opened for business, the company differentiated its approach to content processing by suggesting its method was lower cost, more efficient, and more technologically sophisticated. The competition was caught by surprise. Autonomy then layered on other benefits that drew inquiries from prospects; for example:

- Autonomy had invented a "dynamic reasoning engine" pictured in an early patent as a sophisticated, complex system
- The "black box" in the dynamic reasoning engine could understand content and automatically assign a document to a category even though the specific word for that category did not appear in a source document
- The system was based on advanced statistical procedures, which in the late 1980s, were not widely known nor understood by such search industry leaders as Fulcrum Technologies, IBM, OpenText, Teratext, and Verity.
- Autonomy's engineering allowed distributed processing, painless scaling, and such advanced features as dynamic caching
- The Autonomy DRE and IDOL technologies were immediately applicable to fraud detection, customer support, Web portals, traditional research including business intelligence, content archiving, and dozens of other problem areas with which most businesses struggle on a daily basis.

Autonomy was a competitive firm. Search and content processing companies continue to focus on complex methods used to allow an employee to locate a document. Autonomy offered a black box and strong arguments for the cost effectiveness of Autonomy's sophisticated system and methods. When portals surfaced as a trend, Autonomy offered a Portal-in-a-Box solution. When Endeca introduced Guided Navigation, Autonomy responded with dynamic facets. When visualization caught the attention of senior management, Autonomy introduced visualization outputs for its knowledge and business intelligence versions of IDOL.

Technology, management savvy, marketing, and moxie helped fuel Autonomy's revenue growth. The competitors and analysts comfortable with

kinder, gentler approaches to content were horrified. The annoyance with Autonomy is visible even today (November 2007) in reports produced by CMSWatch, for example.



The inner workings of the Dynamic Reasoning Engine and the “intelligence layer” was partially revealed in US7,272,594, filed in May 2001 and granted in September 2007. The version of the patent illustration has been simplified. The original contains reference numbers to specific paragraphs in the patent document disclosing the methods used by Autonomy to link documents automatically to that a user can access related content discovered by IDOL to be germane to the user’s business task. The patent is important because other mathematical methods are disclosed. These complement the Bayesian methods with which Dr. Lynch is most closely associated.

The systems identified words and a query yielded a list of documents The companies that sell clients profiles about search vendors were quite enthusiastic about Autonomy’s “meaning based computing.” There were thousands of Autonomy customers, and most of them seemed to be satisfied with the Autonomy products and services. The complainers may have been envious of Autonomy’s business success.

Like it or not, by the end of 2006, Autonomy was the most successful search vendor generating revenue from license fees, royalties, services, and savvy management. In 2007, Autonomy was aiming toward $600 million in revenue, and the next closest competitor was approaching $80 million in revenue. With search and retrieval mired in endlessly similar technologies, the

"The bigger the task the more accurate the machine will be over time."--Dr. Michael Lynch, in a story by David Woodward, "Profile: Michael Lynch," Director Magazine, May 2011.

difference boils down to management and marketing. The colorful and aggressive head of Autonomy is Michael Lynch. Compared to the revolving door at most enterprise search companies, Autonomy had stable management, and the company posted sold revenue growth from 2000 to 2006 and an ability to win deals. In the lofty atmosphere of some competitive search vendors and mid-tier consulting firms, Autonomy was different—the company was disruptive and growing at a time when other search vendors were selling out, facing a crisis in revenue, and struggling for a way to package search as a high-value product.

Some History

The founders – Michael Lynch and Richard Gaunt – are from Cambridge University where Autonomy began as a class project.[2] Lynch and Gaunt (and their numerous sales professionals) emphasize that Autonomy's strength lies "in a unique combination of technologies that employs advanced pattern-matching techniques utilizing Bayesian Inference and Claude Shannon's principles of information theory."

Autonomy's technology has its roots in an 18th century Presbyterian minister's mathematics. Bayes's Theorem sets forth a method by which one can derive inferences about what is analyzed. When the Bayesian numerical recipes are applied to information retrieval, the system "learns"; that is, IDOL automatically (autonomously) forms an understanding of the concepts of the processed content.[3] IDOL makes inferences about the information. The method makes "meaning-based computing" work.

$$p(\theta|x) = \frac{p(x|\theta).p(\theta)}{\Sigma_{\theta' \in \Theta} p(x|\theta').p(\theta')}$$

Autonomy's origin was a student research project about Bayesian statistical theory with considerable utility in signals processing.

Michael Lynch, one of Autonomy's founders, said:

> Users are inundated with too much irrelevant information on the Internet. Autonomy's Agentware personalized information solutions utilize Neural Network based Intelligent Agents to dynamically understand user preferences which allow service providers to deliver relevant information. This relationship is part of Autonomy's corporate road map to leverage key partner relationships to offer customers the best solutions.[4]

---

2. The third Autonomy founder is David Tabizel, who graduated from the University of East Anglia and maintains a lower profile than Mr. Lynch.

3. For more information, see "An Essay Towards Solving a Problem in the Doctrine of Chances" at http://www.stat.ucla.edu/history/essay.pdf

Autonomy based its approach to search on math, filed patent applications, and locked up its core methods in a black box. Other search vendors had access to the good minister's math, and focused on adding value via linguistics, semantics, and knowledge bases. Autonomy used these methods as well, but—and the *but* is important—emphasized that its system was "automatic", and Mr. Lynch named the company's name "Autonomy" for a reason. In a head-to-head comparison of search systems, Autonomy was as good or bad as other systems. However, Autonomy's sales and marketing approach focused on:

- Mathematics in the sense that 2+2=4. One can have difficulty arguing with mathematical truths.
- The IDOL system was "automatic," and "automation" was desirable in certain types of labor-intensive work.
- Autonomy bundled Bayes, various technologies, and acquisitions far afield from basic keyword search into a wonderful catchphrase; to wit: "meaning based computing."

Autonomy had math and almost every other search-related function. Instead of licensing a search system and a "semantic" metadata system from Schemalogic or IBM Unicorn, Autonomy offered IDOL plus any other feature a client required.

Autonomy's corporate set up from its early days placed the IDOL technology in a company called Cambridge Neurodynamics. Autonomy then licensed the technology that had to be kept confidential. Autonomy's engineers then used this "black box" to build solutions that incorporated neural network technology, pattern matching, and other sophisticated operation.

Bayes for Beginners: How I Learned to Love Recursion and Stop Counting

Dr. Kate Cowles uses the explanation below in her information science classes at the University of Iowa, and it is a clear statement of the recursive nature of the Bayesian approach to data analysis:

1. Ask a question or pose a problem.

2. Assemble and evaluate the relevant information.

3. Based on current information, design an investigation or experiment to address the question posed in step 1. Consider costs and benefits of the available experiments, including the value of any information they may contain. Recognize that step 6 is coming.

4. Carry out the investigation or experiment.

5. Use the evidence from step 4 to update the previously available information; draw conclusions, if only tentative ones.

6. Repeat steps 3 through 5.

(See http:// www.math.uiowa.edu/ftp/kcowles/s138/lect01.pdf)

Math and Marketing

Autonomy is among the first to say that no set of algorithms can operate as well as a human or be "right" 100 percent of the time. However, Autonomy is adamant about the value of its underlying mathematics, particularly when applied to search and retrieval. When the Autonomy "engine" has abundant data to analyze, it can generate relationships that would be otherwise difficult or too expensive to discover using subject matter experts. As the volume of data to be processed goes up, Autonomy's technology can handle the work. If the source data are "thin" or in some way statistically out of whack, the Autonomy approach needs to be balanced with the same sorts of knowledge bases and word lists that power other enterprise search approaches. IDOL is wrapped with services and interfaces to permit subject matter experts interact with certain IDOL metadata operations. IDOL, however, is off limits for licensee tinkering.

Autonomy packages various advanced technologies into what the company calls its *Adaptive Probabilistic Concept Modelling* (APCM). Autonomy's Technology White Paper references the good reverend, and explain the rock solid mathematical ground on which IDOL stands. Michael Lynch, one of the founders of Autonomy, is a mathematician with a forceful nature. Those challenging him are likely to be met with a mathematical response, not a smile and a pat on the head.

To the chagrin of its competition, Autonomy's approach, therefore, is not a "one size fits all" solution. The IDOL technology is, the company asserts, more for discovery and business intelligence functions as well basic search and retrieval.

"Automatic" and "automation" are the metaphors on which Autonomy's marketing edifice stands. By automatically forming an understanding of the concepts within text or video Autonomy is a more economical system in the long run.

However, the reader should keep several points in mind when trying to sort out the differences among search and retrieval solutions. Each of the enterprise search providers use numerical recipes, including Bayesian methods. If one digs deep enough Bayesian statistics turn up in virtually all of the vendors products. However, search engine vendors fall into different philosophical schools. For example, Delphes, for example, focuses on linguistic methods. Fulcrum, on the other hand, uses keyword indexing and metadata outputs as inputs to ranking procedures. The anti-Bayesian vendors like Siderean and iPhrase put more emphasis on dictionaries, algorithms that attempt to figure out relevance from XML tags, natural language processing, and dozens of other approaches.

On the Autonomy 2001 fall off in revenues: Iain Staples, a technology analyst at ABN Amro, said he had revised his pre-tax profit forecasts for this year to $17m from $29m. Mr Staples added: "The company said only a few weeks ago that it had seen no slowdown in either Europe or the United States. Frankly the management now have a credibility problem. I'm very surprised that they've acted so surprised by this sudden fall-off in sales.--Source: "Dan Sabbagh, "Autonomy Shares Plunge on Profits Alert," Telegraph, April 7, 2001.

Finally, there are vendors who focus on business benefits of a hybrid system like Convera and Endeca.

IDOL is a black box. Autonomy does not provide much detail about the inner workings of its core component. There are clues in Autonomy's marketing literature and white papers.

The basic process involves several steps:

1. A licensee is asked to provide a collection of information or data that is "like" the information the system will automatically process. The collection is called a "training set." Autonomy's indexing system will figure out what is and what is not important and perform the desired indexing and classifying functions automatically. Before deploying the system, Autonomy's engineers, will check and tune the IDOL system.
2. The system indexes content automatically. Periodically, the system administrator or other professional will check to make sure the system is performing within the specification. If necessary, the IDOL system can be retrained. In some installations, no retraining of IDOL is required; for example, scientific and technical content exhibit less drift than business information.
3. Autonomy's throughput can be scaled at any time by adding servers or storage. When the volume of content increases, the licensee can add computing resources. No change to IDOL is required.
4. Because it is built on mathematics and not language, Autonomy allows content to be searched simultaneously in any language and any format, wherever it is stored, and presented with summaries and hyperlinks to similar information, automatically. IDOL can be configured to function as a content repository or as a system that merely indexes, displays outputs, and generates alerts that can be sent to individuals or other systems.

Revenue Growth

Autonomy was founded by Dr. Michael Lynch and Dr. Richard Gaunt. A third founder was David Tabizel, an acquaintance of Dr. Lynch's. The table below provides a summary of Autonomy's gross revenue and pre-tax profit from 1996 to 2007. Some financial reports are stated in pounds and others in US dollars. Due to the restatement of Autonomy's finances and the com-

pany's shift in accounting methods, the table reflects various calculations to deal with these issues.

Table 2: Autonomy Financial Snapshot, 1996-2007

	Revenue	Profit/Loss
1996	242,391	(4,038,455)
1997	2,577,380	(2,158,250)
1998	8,300,000	(2,645,108)
1999	16,511,000	(678,000)
2000	65,422,000	14,545,000
2001	52,600,000	9,366,000
2002	50,961,000	6,518,000
2003	54,881,000	7,680,000
2004	64,765,000	7,683,000
2005	96,032,000	12,628,000
2006	250,682,000	56,319,000
2007	343,409,000	91,447,000

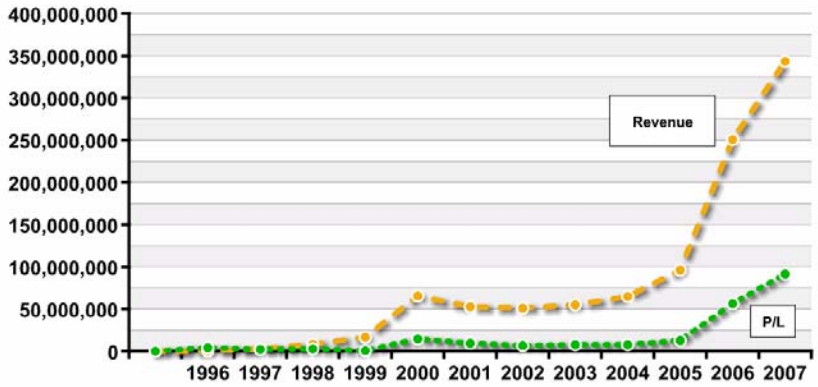
Several observations are warranted:

1 Between 1999 and 2000, Autonomy pulled off a remarkable sales achievement. The company's revenues surged from $16 million to more than $65 million, a feat no other search vendor had previously been able to achieve; that is, growth of more than 400 percent and reversing a loss of nearly $700,000 into a surplus of $14.5 million.

2 Between 2005 and 2006, Autonomy grew revenues from $96 million to more than $250 million and boosting its profit from $12.6 million to more than $56 million. At this time, Endeca was roughly one-tenth the size of Autonomy and Fast Search & Transfer was stalled at about $140 million and scrambling to deal with management and financial challenges.

'This latest substantial increase of our reach into the telecoms market demonstrates excellent progress and a continuing validation of both our technology and our business model. Our increased penetration into various industry sectors highlights the fact that Autonomy's infrastructure technology serves as the supporting layer for the automation of unstructured information across all applications, and is therefore a must-have technology in the drive to increase business efficiency.'—Dr. Michael Lynch in 2001 after announcing deals with telecommunication companies. See http://www.pennypinchers.org.uk/password/pennydrop/autonomy.htm

Viewed over the long term, Autonomy's financial performance looks like this:



What is clear is that Autonomy began to grow in 2005. The shift at Autonomy was a realization that large acquisitions, aggressive sales, and cost management were the ingredients of financial success. In 2005, Autonomy was nine years old. Unlike iPhrase which hit a financial wall after six years in business, Autonomy dealt with the "ceiling" on organic growth from IDOL technology via acquisitions.

An important lesson can be drawn from the revenue, profit/loss, and trajectory of Autonomy through time; namely, in order to build a robust information retrieval, Autonomy had to expand into adjacent market sectors. Autonomy, therefore, began walking a route that would lead the firm to more acquisitors, more aggressive sales tactics, and more stringent cost controls.

Endeca, a company founded at about the same time, did not emulate Autonomy. In 2007, Endeca, a privately-held company, was roughly one-tenth the size of Autonomy with revenues of about $50 million per year. Dr. Lynch had found the turbo-charger for enterprise software growth. Autonomy was not a search vendor. Autonomy sold infrastructure and solutions to problems associated with digital data and information.

Snapshot of Michael Lynch

Debate.org provides a biographical sketch of Dr. Lynch. Debate.org reports:

> Michael Richard Lynch OBE is the co-founder of Autonomy Corporation. His entrepreneurship is associated with Silicon Fen. Autonomy is a leader in the area of computer understanding of unstructured information, an area which is becoming known as meaning-based computing. In October 2011, Autonomy was sold to Hewlett Packard for $11bn, making Lynch a dollar billionaire. He was born in Carrick-on-Suir, County Tipperary, Ireland, in 1965, but his family moved to England, where he grew up near Chelms-

"We like to make a big bang."—Richard Gaunt, Peggy Waldman, "Chief Technology Officers Talk Technology, "Sybase Magazine, Winter 2001, page 21

ford, Essex. His mother was a nurse and his father a fireman. In 1976, aged 11, he won a scholarship to Bancroft's School, Woodford. From there he went to Christ's College, Cambridge to study natural sciences. He combined mathematics, biological and physical sciences, taking the combination of advanced physics, mathematics and biochemistry in the IB Tripos. For part II, he chose electrical sciences where he first met Dr. Peter Rayner, his mentor in the signal processing laboratory of the engineering department.

Three points warrant comment:

1 Dr. Lynch comes from the UK's middle class. Is it possible that his management style and entrepreneurial passion are in part due to his having to rise in the class conscious and class sensitive environment of Cambridge University in the 1980s?

1 Dr. Lynch has a multi-disciplinary background. His exposure to biology, chemistry, physics, and mathematics. Compared to some search and content processing entrepreneurs, Dr. Lynch approaches the challenges of digital information from a different angle than a search engine programmer. Is the multi-disciplinary background partly responsible for his comfort with the ambiguities of signal processing and Bayesian-Laplace methods?

1 Dr. Lynch's mentor is Dr. Peter Rayner. Did Mr. Lynch's work with Dr. Rayner provide the eager Ph.D. student with encouragement to embrace methods associated with cryptoanalysis, advanced numerical procedures involving a range of comparatively new methods, and a solid foundation in Dr. Shannon's information theories?

Autonomy is interesting because it is the first vendor to base search and content processing on methods that once were classified and not widely applied to problems in enterprise information analysis.

My view is that the critics of Autonomy lack the mathematical foundation to understand Dr. Lynch's development of IDOL from the early 1990s to mid-2007.

Other Executives

As the company moved through its different owners, not surprisingly, top management changed. The original founders were:

- Richard Gaunt, a graduate of Natal University, co-founder and technical director until 2005. He was instrumental in the digital recording technology in Dremedia (a unit of Autonomy) as well as IDOL
- David Tabizel, co-founder, source of funding for Autonomy in 1995
- Sushovan Hussain, chief financial officer Chairman New Jersey Autonomy Inc., a foreign for profit corporation in Florida. Mr. Hussain joined

"Autonomy will allow the Department of Labor to save time and money because it minimizes the need to have employees and Disability Direct partners manually categorize, tag and insert hypertext links between related content,' said Richard Gaunt, Autonomy chief technology officer and co-founder. ``For example, as new information is created by a local government, our software analyzes the main concepts and then automatically categorizes it and inserts links to related content in real-time."--- Richard Gaunt, Autonomy founder and CTO. Source: http://biz.yahoo.com/prnews/010626/nytu029.html

Autonomy in June 2001. This unit was set up in December 2005. Sushovan Hussain, 40, has served as our Chief Financial Officer since June 2001 and was appointed a Director in June 2003. Prior to joining Autonomy, Mr. Hussain worked for LASMO plc, one of the world's largest independent oil and gas exploration companies, where he held a number of senior international financial positions, including three years in the Corporate Development department, charged with acquisitions and divestments. Mr. Hussain received his BA in Economics from Cambridge University, England, and became a qualified Chartered Accountant while employed at Ernst & Young in London. Mr Hussain is also a non-executive director of Zynap Limited for which he receives no additional remuneration. Mr. Hussain was appointed to the Board in June 2003.

- Stouffer Egan, chief executive officer US, Stouffer Egan, US CEO in 2004. Mr. Egan joined Autonomy in March 2001 as Director of Global Accounts, and has served as Chief Executive Officer, U.S. since 2002. Prior to joining Autonomy, from February 2000 through March 2001, Mr. Egan was Vice President of Corporate Development for Leading Side Inc. From January 1995 through February 2000, Mr. Egan held various positions with Dataware Technologies, most recently as Vice President, Corporate Development. Mr. Egan holds a B.A. in Economics from Trinity College, CT.He is the spark plug for Autonomy's sales method. (David Appelbaum, was the US CEO in 1999.)
- Anthony Bettencourt, US president of Autonomy. He was the CEO of Verity at the time of the Autonomy buy out
- Andrew Kanter, an attorney who served as Autonomy's chief operating office. Although slightly more mellow than Dr. Lynch, Mr. Kanter is firm, tenacious, and willing to confront those who criticize Autonomy.

Autonomy may have as many 300 professionals engaged in research and development. The company's chief technology officers are:

- Dr. Peter Menell, Autonomy CTO, a neuro physiologist. Dr. Peter Menell, chief technology officer. Ph.D joined Autonomy's Engineering and Technology Solutions unit in 1998 and has served as Chief Technology Officer since 2004. Prior to joining Autonomy Dr Menell conducted computational and neuro-physiology research in visual and auditory impairment.Dr. Menell holds a BA (Honors) and Master of Science from York University and a Ph.D. from Oxford.
- Elroy Avila, US CTO, electric vehicle research.

"Unlike limited keyword-based technologies, Autonomy automatically analyzes, identifies and prioritizes the main concepts within any piece of content, enabling customers in all verticals to automate a broad range of business critical tasks previously dependent upon lengthy, manually intensive processes."—Autonomy Annual Report for 2004

Selected Clients

Autonomy states that it has more than 20,000 customers. No complete list of the company's licensees is currently available. Selefte4d Autonomy customers include:

- Accenture
- AstraZeneca
- BioQ
- Boeing
- Butterworth's (LexisNexis, a unit of Reed Elsevier)
- Department of Homeland Security
- Deutsche Bank (also an Autonomy stakeholder)
- Ferrari
- General Motors
- HSBC
- IBM
- Lexmark
- London Metropolitan Police
- New York Stock Exchange
- Philips
- Royal & Sun Alliance
- Shell Oil Company
- Sun Microsystems
- US Department of Energy
- Vodaphone.

The OEM Deals. Autonomy acquired Verity and obtained a revenue stream from third-parties who licensed the Verity search system for information retrieval. Autonomy added Verity's OEM (original equipment manufacturing) deals to its own and become the largest licensor of search technology in the world.

'Autonomy's IDOL Server connects people to content, content to content, and people to people through modules that enable organizations to integrate various personalization, collaboration, and retrieval features. The server provides a knowledge repository... The server seamlessly extends and integrates with the company's e-business suite, allowing rapid deploy-ment applications that span the enterprise and leverage AI-assisted technology to harvest knowledge assets."--Knowledge Management textbook turban_tutor05_W183-W199-hr

Autonomy OEM Licensees (Selected)

OEM Licensee	Key Technology
Adobe	Consumer and commercial software
Cisco	Network hardware and services
Citrix	Enterprise and consumer remote access systems
EMC	Storage
Hewlett Packard	Also used by EDS, an HP subsidiary
IBM	Enterprise solutions, services, and products
Iron Mountain	Archiving and e-discovery services
Kana	Customer support systems
Matrix One	Collaboration solutions
Novell	Enterprise infrastructure
Openwave	Mobile solutions
Oracle	Also used the Oracle subsidiary Hyperion
Support Soft	Enterprise software
Sybase	Database and data management software
Symantec	Enterprise software
Tibco	Enterprise infrastructure
Verdasys	Archiving and security services
Xerox	Enterprise systems, hardware and services

Selected Partners and Integrators. From 1996 forward, Autonomy sought partners and resellers. The tie ups provided a flow of sales leads and a mechanism for delivering additional services to IDOL licensees. In 2005, when Autonomy purchased Verity, Autonomy inherited a consulting and services arm. For-fee engineering services were added to the Autonomy menu of products after 2005. Autonomy continued to acquire promising integrators and resellers in order to secure this revenue stream.

Table 3: Autonomy Technology Partners (Selected)

OEM Licensee	Key Technology
Accenture	Management and technology consulting
Boeing	Manufacturing
CSC (Computer Sciences)	Infrastructure and services
Canon	Industrial and consumer products
Capax Global	Financial services
Capgemini	Management and technology consulting
Captaris (now a unit of OpenText)	Document management
Documentum (now a unit of EMC)	Document management
Dow Jones & Co.	Publishing
Fujitsu	Manufacturing and infrastructure

Hewlett Packard	Hardware, software, and services for consumers, and commercial entities
IBM	Services, hardware, and software
LexisNexis (a unit of Reed Elsevier)	Publishing
Lockheed Martin	Infrastructure services and manufacturing
Logica	Business and technology services firm
Morse (a unit of 2e2)	Information technology services
Northrop Grumman	Infrastructure services and manufacturing
SAIC	Infrastructure and services

Selected Acquisitions

A key part of Autonomy's financial performance depends on its acquisitions. Autonomy does a good job of upselling acquisition's customers. Autonomy also brings financial and sales management practices to the acquired companies.

Table 4: Selected Autonomy Acquisitions

Year	Company	Est. Amount Paid	Comment
1999	Intracon	n.a.	Autonomy bought a 33 percent stake in this distribution partner
2000	Nordic	n.a.	An early Autonomy reseller serving Scandinavia
2000	SoftSound	$800,000 (est)[a]	A company involved in speech recognition. Autonomy acquired patents which were assigned to Longsand Limited, an Autonomy entity
2001	Dremedia	n.a.	An entity set up to commercialize the Drs. Gaunt and Lynch technology developed by Cambridge Neurodynamics
2003	Virage	$14 million	Automated audio and video surveillance
2005	Global Linxs	n.a.	An integrator and developer of a patent information system. The founder, Jakob Riegger, managed Autonomy's consulting services
2005	eTalk	$72 million	Call center monitoring software
2005	Verity	$08 million	Fading search vendor with a large OEM and enterprise customer list. Verity itself had acquired Cardiff (Liquid Office), Ultraseek (developed by InfoSeek, sold to Inktomi, and then Yahoo bought Inktomi and sold Ultraseek to Verity), 80-20, Native Minds, Dralasoft, Keyview, and 64K. Cardiff had acquired AudienceOne.
2006	OpenV China Holdings	n.a.	Set up to facilitate deals in China
2007	Blinxx	n.a.	Blinxx was founded by a former Autonomy executive, Blinxx was repurchased by Autonomy and then taken public in a $250 million in 2007
2007	Meridio	$45 million	Records management system
2007	Cambridge Neurodynamics	n.a.	Autonomy purchased the balance of the 1991 company founded by Drs. Gaunt and Lynch

Table 4: Selected Autonomy Acquisitions

Year	Company	Est. Amount Paid	Comment
2007	Zantaz	$375 million	Email archiving and cloud services. Zantaz had acquired Single-cast, SteelPoint, and Educom TS.
2007	Fast Search	n.a.	Autonomy purchased the Fast Search US government sales team
2007	Promote Multichannel Technology	n.a.	A Cambridge, UK-based social media company

a.Source: http://boards.fool.co.uk/softsound-purchase-5981310.aspx

Does Autonomy preserve the brand identify of the companies and products it acquired? The answer is, “It depends.” Ultraseek is a lower-cost search option, but the ties back to Inktomi and InfoSeek are not evident. The Verity K2 brand was initially “supported” and then supplanted for new sales resulting from Verity prospects. The identity of Virage persists, but Autonomy generates different brands which may offer products developed by acquired companies; for example, Audentify and Aungate. Autonomy is an extremely adept marketer of its technology. Over time, acquired products are positioned as IDOL, the Autonomy ultra-secret “black box.” The Autonomy white papers show functions plugged into IDOL. The Zantaz deal put Autonomy in the cloud computing market. But below the surface, the integration is usually “good enough.” A licensee with appropriate expertise and resources can develop a system that delivers the desired functionality. The clever Portal-in-a-Box product is less of an appliance and more like an high-end erector set or Lego set.

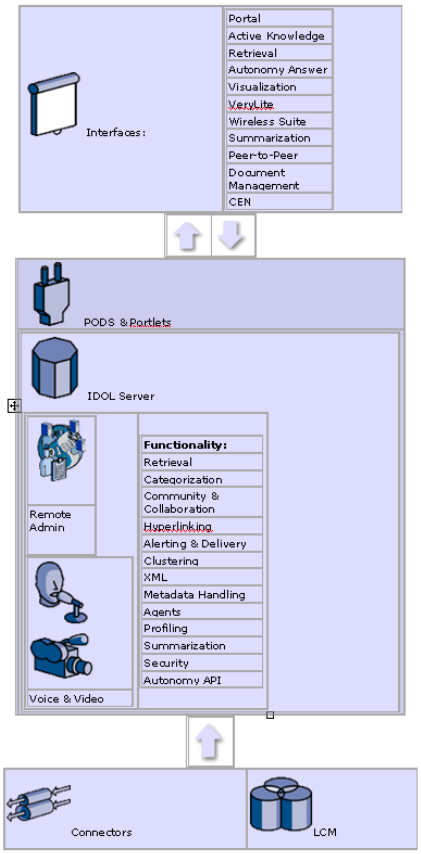
A Schematic of IDOL

The Autonomy IDOL system is positioned as “automatic.” The diagram below provides an overview of the principal components of an enterprise search set up. Enterprise content processing architectures are, on first glace, almost indistinguishable.

The architecture of IDOL is no exception. Where IDOL differentiates itself is the mathematics of signal processing. Since 1996, Autonomy has add to the IDOL system features and functions ranging from support for languages other than English, a user interface to make it easier to update word lists, and other components. I think of these enhancements and responses to competitors’ innovations as plug ins. When a competitor adds text analytics visualizations, Autonomy responds quickly. The IDOL platform is easily extensible both for licensees, resellers, and for Autonomy’s research team.

The agility of Autonomy is due in part to the “black box” approach. The DRE operates with other modules and components in prescribed ways.

When a competitor offers eDiscovery functions, Autonomy either adds those functions or assembles modules to deliver similar functionality.



The interface layer can be tuned to meet the specific needs of the user in a particular unit of the licensing organization. Interfaces are available for most business processes.

The "core" of Autonomy's platform is IDOL. Note that IDOL has components to support security and rich content. In addition, there are subsystems within IDOL that apply the Bayesian and other algorithmic methods to information processing, query processing, indexing, value-added tagging, entity extraction. Note that the system can work with market up "information objects" in a way similar to methods in XML data management systems.

The connector layer allows IDOL to access more than 400 content formats. A library content management system is included.

The lower layer represents the structured, unstructured, and semi-structured data available to IDOL.

In terms of marketing, the add ons and modules do not change the DRE core. Because of this architectural approach, Autonomy can quickly react to competitors' innovations. Not surprisingly, Autonomy is viewed as a company that takes no prisoners in a competitive situation.

The Autonomy architecture consists of a handful of building blocks. These are:

1 Content acquisition and processing. IDOL ships with filters for more than 200 file types. There are also connectors to the document repository. When a user requires content, the content is served from the XML repository. Autonomy supports rich media.

"Autonomy and Cognos share many common global enterprises and government agencies as customers today, (May 2006). This integration opens new markets for both Autonomy and Cognos. All of a sudden, the 'meaning' of thousands of customer communications that might otherwise sit in a database are understood and become actionable in the business intelligence software."—Stouffer Egan, Autonomy. Source: "Autonomy and Cognos Form a Partnership, May 15, 2006 via the Beye Network.

2 Indexes or representations of processed information. Processed content becomes metadata, word pairs, and tables of probabilities. Think of these constructs as collections of content objects with unique patterns. When a query is converted to a pattern, Autonomy employs pattern matching, weights, values, and probabilities to retrieve relevant content. These indexes and representations are hierarchical so a lower branch of a hierarchy is automatically known to be a branch of a higher level concept. Thus, classification and clustering are available during query processing.

3 The search module matches patterns and taps the representations for categories and classification metadata.

4 Administration. Autonomy provides configuration files that licensees can edit. With each new release of IDOL, Autonomy has expanded and added graphical interfaces, dashboards, and control panels.

The Digital Reasoning Engine is embedded in the IDOL system. The black box of algorithms calculate the probabilistic relationship between multiple variables and determine the extent to which one variable impacts another. Autonomy's software attempts identifies a context of a piece of unstructured information using the metadata, the hierarchies, and the representations of processed content. Autonomy has added operations that make use of the registered user's work role and previous searches.

Autonomy's approach in the Digital Reasoning Engine operates regardless of the language of the source content. The idea in signal processing is to discard that which is redundant and carries a small payload of meaning. Autonomy has added language-specific dictionary support and other linguistic features to support online translation.

Overall Search Features

Autonomy offers licensees several packages of search software. The principal ones are:

- Autonomy Server (IDOL)
- The application builder consisting of toolkits for developers
- An updater with spiders and supporting software to discover new or changed data available to the system
- Optional video and audio indexing and retrieval packages.

The IDOL system:

- Finds pages conceptually similar to user searches and other documents
- Summarizes and extracts relevant content from processed content

- Supports keyword searches, Boolean searches natural language, and “more like this” queries
- Factors in the searcher’s context when determining relevancy and clustering content
- Indexes structured data for field searching.
- Supports a thesaurus for control of synonym searching.
- Displays related documents

Autonomy offers a number of ways for a user of IDOL to find information. These include:

- Agent-based search that retrieves information automatically based on what the user is typing into Word or some other application
- Support for Boolean queries
- Conceptual search; that is, a broad term of any type
- Directed navigation (a version of Endeca’s Guided Navigation)
- Free text query
- Fielded query
- Alerts of new content matching a profile
- Graphic displays which show clusters or groups of content related to a topic, concept, or term
- A Google-style laundry list of documents for the user to browse.

A stream or channel of content on a topic.

Licensees, at their option, can obtain specialized routines that perform additional functions in the IDOL environment. Among the options available are:

- Java applet to visualize information domains
- High-performance systems of hardware and software that support near-real time content processing
- Targeted advertising and promotions (public and commerce sites) that is roughly congruent to the metrics and feedback functions in iPhrase and InQuira systems
- Alerts to users when new content relevant to their information needs enters the system
- Expert identification and collaboration features.

Federating Content from Multiple Repositories

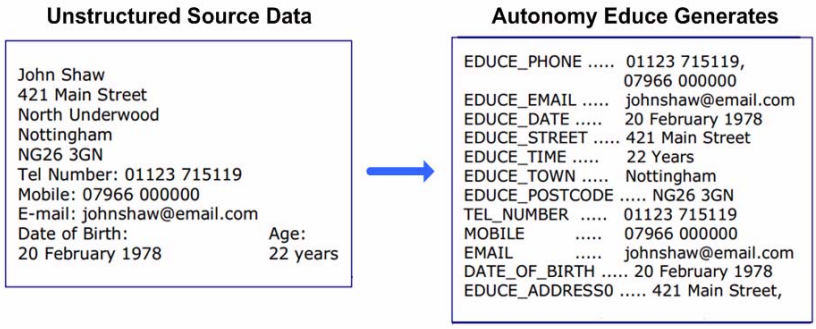
For enterprise search Autonomy recognizes that information is typically stored in different formats and file systems. Customer records are held in a CRM system, catalog information in an Oracle database, and email in a

"Autonomy enlisted in war on terror."— Silicon.com, October 21, 2002

specialized mail server archive. For an individual to perform a business task efficiently, each authorized user requires access any relevant information in the organization.

Searching Unstructured and Structured Content

Autonomy provides an "Education" module. This subsystem extracts key information from the content transformed to IDOL XML. For example, a resume listing an applicant's name, address, skills set, companies worked for, and job title sought, written in Word format with zero metadata, can be transformed into a structured document using the Education module.



Education will read the document and automatically populate metadata fields by educing their correct content from the unstructured text. Once this has happened the document can be assembled as an XML file with the NAME, ADDRESS, SKILLSET, PREVIOUS COMPANIES, and POSITION fields completed by Education.

In addition, Autonomy's ability to exploit the structure inherent or implicit in information enables applications such as parametric retrieval and guided navigation, where an understanding of how one metadata constraint relates to multiple others is required. For applications where content is highly normalized such as commerce, supply chain, ERP etc., Autonomy provides parametric capabilities to enable a drill down and navigation process that guides the user to the right answer with a minimum of effort.

Autonomy also allows a user or other system to submit a query that is like a formal Structured Query Language instruction to select a mobile phone plan that costs less than $100 per month, offers SMS and Web browsing services, and has no early termination fee. The first two constructs are present in the database and can be handled with parametric retrieval; the third is buried in a contract in unstructured textual format, and can only be retrieved via conceptual understanding of the query.

Autonomy's query processing system will return an "answer" to the user or the software component submitting the query.

Autonomy is the first enterprise content processing system to support mixing and matching different search methods within the same query without requiring specialized syntax.

Configurability

Other modifications are supported via the IDOL API. The Autonomy toolkit allows licensees to customize Autonomy's modules according to their specific requirements.

The Application Programming Interfaces (APIs) allow licensees to implement only the functionality necessary for a specific application. Vignette, for example, uses Autonomy to provide basic search and retrieval within the Vignette content repository. Licensees receive the APIs and sample code with the Autonomy Server.

Autonomy provides the following features for optimizing a licensee's system:

1 **Stop word lists.** These are lists of *empty words* that do not carry any significant meaning. In grammatical terms these would normally be prepositions, conjunctions, and auxiliary verbs. For example, in English words such as *the*, *a*, and *to* are often considered stop words. These words can be safely ignored when Autonomy's system is processing content.

2 **Stemming.** In most languages certain variations of a word can be stripped to obtain the main stem of the word. In English for example, the words criminal and criminologist can all be stripped down to its stem *crim* without much loss of meaning in Autonomy's system. Stemming rules can be safely used when processing text in order to obtain a list of unique words. Note that Autonomy does not require the use of stop lists or stemming rules. Autonomy's statistical analyses would normally determine the importance and relationship of those words. However, an initial configuration of a stop list and stemming rules allows the IDOL

The 1998 "penguin" understanding example: "A keyword-based search, on the word 'Penguin', may provide the user with a volume of irrelevant items across the corporate intranet that includes the word Penguin. On the other hand, using a 'intelligent' approach, the computer understands the context of a specific search and would know that you are looking for information about Penguin chocolate biscuits and not the bird."—Richard Gaunt, co-founder, Autonomy, "the Hidden Cost of Knowledge Management," Inside Knowledge, September 1998.

Server to ignore empty words and treat a set of words as one so that storage resources and processing time can be reduced.

```
client aci = new client();  // aci Client Object
aciObject acioDRE = aci.aciObjectCreate(aciObject.ACI_CONNECTION); //
Create DRE connection object
acioDRE.paramSetString(aciObject.ACI_HOSTNAME, "192.168.14.51");
acioDRE.paramSetString(aciObject.ACI_PORTNUMBER, "3002");
acioDRE.paramSetInt(aciObject.ACI_CONN_TIMEOUT, 60000);

aciObject acioCommand = aci.aciObjectCreate(aciObject.ACI_COMMAND); //
Create query command object
acioCommand.paramSetString(aciObject.ACI_CON_COMMAND,"Query");// Set
action

acioCommand.paramSetString("text", keyword);
//acioCommand.paramSetString("FieldText",prevQuery);

//acioCommand.paramSetString("DatabaseMatch",classifiedDB);
acioCommand.paramSetInt("MaxResults",300);
acioCommand.paramSetInt("Characters",3000);
acioCommand.paramSetBool("Spellcheck",true);
acioCommand.paramSetString("Combine","Simple");
//comment out IgnoreSpecials=true to enable Boolean search
//acioCommand.paramSetString("IgnoreSpecials","true");
acioCommand.paramSetString("Sort","DISTRICT:reversealphabetical");

aciObject acioResult1 = acioDRE.aciObjectExecute(acioCommand1);
```

3 **Editor interface.** Autonomy provides a manual editor that allows users to edit terms and their relationships and weights. However, for accuracy and ease of use, Autonomy prefers that licensees use Autonomy's "editing by example" where documents can be used as examples of things to be included or excluded from categories. Autonomy's approach to training documents is that a corpus is used to "teach" the system about words, phrases, and documents. The manual editing is similar to assigning a weight to an identified term or adding a new term and manually assigning the term a weight.

Content Processing

Autonomy's robust platform contains a number of content processing and information retrieval capabilities. In 2002, throughput on a basic Autonomy IDOL installation was two to three gigabytes per hour.[5] Particular functions that are noteworthy include:

Concept Identification and Relationship Matching

The IDOL system ingests an "information object"; for example, e-mail or a Word file. The system autonomously locates related information during

5. Data are from Autonomy's performance white paper, 2002.

query processing. Autonomy uses hierarchies of metadata. These are updated as new content is processed by the system. Metadata, concepts, models and hierarchies can be generated automatically. For more rapid deployment, the licensee can make word lists, dictionaries, taxonomies, and other knowledge bases available to the Dynamic Reasoning Engine.

This 2005 interface shows Autonomy IDOL as a browser-based search system. Notice that the choices have been narrowed for the user.

The Summarization Function

When a user is confronted with a number of relevant hits to longer documents, IDOL can generate and display a summary of a relevant document. The system can also identify, extract, and display the most relevant snippet of a source document.

Context Services

Users and documents have a context; that is, each exists in a work task or in an organizational unit, topic area, or some other “space.” IDOL can generate a result list or point to content that matches a particular user’s information need related to a work task. Furthermore, the context function operates for each query approach.

One feature of the monitoring method invented by Autonomy is that the IDOL Server notes what a user does with content. This information is added to the user’s profile and used to understand user interests. The monitoring agent identifies changes to those interests and updates the user’s profile. IDOL can automatically identify affinity groups within a user community. A pharmaceutical company with worldwide operations can easily identify researchers at different locations sharing a particular interest or line of inquiry.

Categorizing and Clustering

IDOL processes search results and groups them by relatedness. Vivisimo introduced a similar function years after Autonomy introduced the approach. The clustering function operates automatically and licensees can tune the clustering engine to meet the specific needs of a user or particular group of users. Clustering permits point-and-click exploration of a group of hyper-links to documents that could be otherwise be overlooked in a relevance ranked results list.

IDOL can generate and update its list of categories. Documents have multiple metadata tags related to categories. The alerting function makes use of the categorizing and clustering functions.

The functions of the clustering and categorization subsystem can be tuned via a configuration file. Scripts control more fine grained user-configurable settings.

Smart Agents

Autonomy IDOL makes use of software agents. These intelligent functions note tags, index terms, and entities in processed content. When a watched item or tag is noted by the agent, IDOL can perform operations on that e-mail, document, or video. The smart agents note patterns in the processed content, thus performing a high-level function as well as a “watch list” operation. The method makes use of the probabilistic “scores” that IDOL generates when content is processed. When IDOL is tuned, processed content and the probabilistic scores are used to “learn” about new content in a collection.

Other Text Processing Operations

IDOL includes a number of text processing capabilities. These include:

"Autonomy outperforms all vendors on the technology front... scoring the category maximum in every assessment. In three categories; namely search and query capabilities, visualization and navigation capabilities, and interoperability and integration; the solutions functionality is deemed to be the best-in-class.--Autonomy's "Intelligent Universal Search," page 2.

- Automatic title generation/extraction for a document
- Creating logical fields by aggregating existing fields within the document
- Decoding and transforming base64 encoded data
- Deduplicating characters and identical documents
- Extracting tags and data from XML files
- Filtering rules configured to process a client's source content; for example, file size, file type, etc. IDOL offers a filter tool of ready-to-use parametric operations. The software components can extract information from source data
- Generating a checksum for each document
- Identifying data, extracting it, and tagging it from structured and unstructured information
- Inserting hyperlinks to source documents and related documents
- Recognizing and extracting digits such as a social security number and inserting the number in an appropriately tagged field
- Remapping fields
- Removing numbers in a document or corpus
- Removing specified characters
- Rendering of proprietary file formats such as PDF, Excel or PDF into HTML which can then be moved to a directory. This allows the document to be served up as HTML to make viewing the content easier.
- Replacing a specified string with a blank
- Specifying at which character a field should end
- Specifying at which character extraction begins
- Splitting documents into two or more sections, paragraphs, etc.
- Stripping HTML to yield ASCII text.

Security

Autonomy provides a range of security features. These span security management, including front-end user authentication, back-end entitlement checking and secure encrypted communication between the IDOL Server and its client applications, with 128-bit Block Tiny Encryption Algorithm or BTEA. Autonomy asserts that its "mapped" security model scales in an enterprise.

IDOL is modular; therefore, communications among its processes can be encrypted via Secure Sockets Layer. The assumption is that methods do not exist to intercepting the content of traffic among IDOL modules.

Autonomy emphasizes that its modules can operate in a secure communications mode with 128-bit encryption. Other Autonomy security features include:

- Granular Privileges LDAP Authentication Pluggable Authentication Versioning
- Content Approval
- Audit Trail
- Sandbox
- Login History
- Session Management SMB Authentication NTLM Authentication NIS Authentication Kerberos Authentication Problem Notification SSL Compatible
- SSL Logins
- SSL Pages
- Captcha
- Email Verification

Real-Time Processing

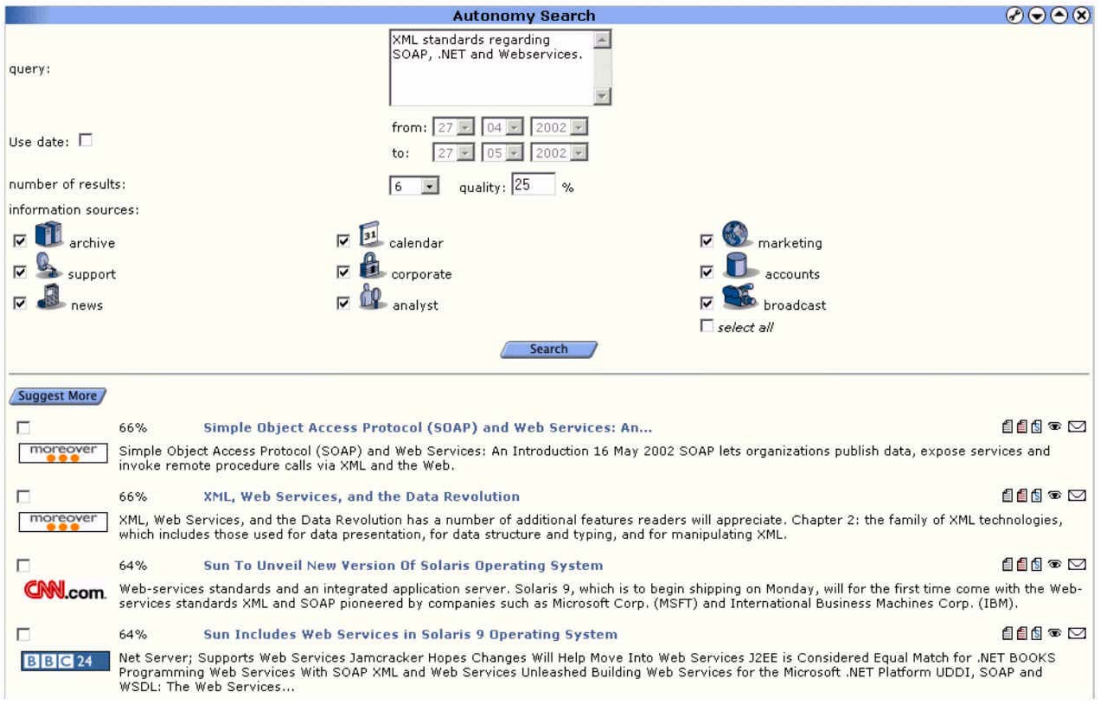
Autonomy has claimed “real time” content processing since 1996. “Real time” is rarely defined by enterprise search vendors. As the volume of data to be processed goes up, the costs of deploying a system that can keep pace with millions or hundreds of millions of new and changed documents rises as well. Autonomy has invested in engineering to reduce the latency in content and query processing.

Autonomy’s operations occur in a low-latency setting when appropriately resourced, provisioned, and configured. In practical terms, the content processed by the system becomes available to users and to other IDOL processes within minutes of its entering the IDOL pipeline.

Systems that deliver low latency response times required for stock trading, the licensee has to put the bandwidth, network hardware, and computer hardware in place. Financial trading firms can afford to make an investment in near-real time systems that exhibit millisecond latency.

Since real time is not defined, most organizations assume that its implementation of any content processing system operates in real time. In the early 2000s, Autonomy could process content at a rate of gigabytes per hour. The number of gigabytes processed varies from implementation to timplementa-

tion. Autonomy's partners can package high-performance gear with the IDOL system if the licensee requests this type of solution. Standard enterprise systems often require upgrades in order to maintain the performance of a system like IDOL.



The user interface for near real-time news access. This interface dates from 2002 and displays options for the person looking for information.

I have observed query latency in the four minute range. The problem was the licensee's infrastructure and on premises configuration, not the Autonomy system. Visualizations for business intelligence applications or alerts for fraud detection applications are available, and these also require appropriate resources to work in a low latency way. The time required to render visualizations depends upon the nature of the query, the specific visualization required, and the number of data points plotted.

Rich media places significant stress on hardware and storage systems. High-performance gear and on-going optimization are necessary when dealing with audio and video content. The size of the files demands super-computer performance. IDOL takes advantage of caching, parallel processing, and other engineering methods to reduce latency in the IDOL system.

Business Intelligence

Autonomy was among the first of the search vendors to embrace business intelligence. An organization can look at Autonomy's approach as a business intelligence framework. What set Autonomy apart when it first pushed into fraud detection and text processing for the intelligence community was IDOL's ability to detect patterns. In the late 1990s, search meant entering two or three terms in a search box and working tediously through a relevance-ranked list of documents.

Content hot spots can be explored with a mouse click. The Autonomy IDOL platform provides a number of visualization options, and the system can be extended with third-party visualization tools if IDOL is used in an environment with an i2 Ltd.-type system.

Autonomy's success in business intelligence and text mining is one consequence of the system's ability to handle large-scale content streams in near-real time. Autonomy's tools like its Application Builder Toolkit allows the licensee to tailor processing operations to highly particular content processing recruitments such as those in intelligence and financial services deployments

Not Autonomy.

The core technology uses pattern recognition to find relevant words and related concepts. Users can express queries without having to know the exact words to snag the information needed. Autonomy's system can be used to make a faceted search interface available. Its outputs can be manipulated

by other Autonomy modules or piped into third-party applications for discovery analysis.

Autonomy's platform can manipulate rich media; for example, voice conversations. IDOL converts speech to text and then processes the output. A human does not have to listen to conversations which is time consuming and expensive. An analyst can use the Autonomy business intelligence interfaces to identify concepts or probe into specific content flagged in a text or graphic display.

Autonomy has created an "investigation manager" module for government investigators. The system permits near-real time monitoring and analyzing of person-to-person communications. Autonomy differentiates by offering the IDOL technology as part of a comprehensive information and intelligence platform.

As a student at Cambridge, Mr. Lynch understood that Bayesian statistics could have profound implications for systems attempting to wriggle meaning from unstructured data. Neurodynamics' software and systems—the core technologies in today's Autonomy—are outgrowths of Mr. Lynch's probabilistic modelling and digital signal processing technologies developed.

In the early 1990s, commercial search and retrieval systems required that users know exactly how to phrase a query to get information about a topic. Intelligence analysts took one look at the outputs of a Dialog or LexisNexis system and concluded:

1 If we knew what we were looking for, then we would be able to make the Boolean systems provide information. A query on the Dialog system would look like `SS (strang* OR chok* OR garrot*) AND (Smith* OR Smyth*) AND UD=9999`

2 The information manipulated by intelligence professionals and police was usually not in commercial databases. The data were in the form of ASCII notes typed by a case officer into a terminal, newsfeeds from various services with little formatting in common between Agency France Press and Pravda, or from different electronic data obtained from credit card companies, banks, and intercepts.

Mr. Lynch found a ready market, first in the U.K. and then in the U.S. Even today, Autonomy is viewed as the leader in text mining technologies in many intelligence entities.

ArnoldIT Opinion

Since 1996, Autonomy has been a major factor in the enterprise search market. In the last decade, Autonomy has been quick to jump on trends that bloomed in the scorched earth some enterprise software vendors left behind. Large-scale system failures created an appetite for "portals." Autonomy responded with its "portal in a box" product. Then knowledge management appeared to provide "answers" that keyword search could not. Autonomy promptly positioned its product as delivering automated knowledge management. When organizations realized that unmanaged digital content was an increasingly difficult problem, Autonomy purchased Interwoven and integrated Web content management into IDOL's repository system. When arch rival Verity faced softening revenues, Autonomy's effective marketing, and end-of-life technology, Autonomy acquired the company and offered to support Topic and K2 installations while adding direct support of K2 indexes to IDOL. Autonomy moved with equal alacrity into fraud detection, analytic services for police and intelligence agencies, and predictive analytics based on real time content processing for financial institutions. Autonomy embraced the cloud and the need for secure content archiving with its purchase Zantaz. For a decade has been quick to capitalize on its meaning-based computing catchphrase and its closely guarded technology.

Critics of Autonomy point to the company's aggressive sales tactics. Some licensees have balked at the cost of implementing a comprehensive IDOL system. Competitors make caustic comments that Autonomy takes their best ideas and "glues" them on the IDOL system. Search consultants walk on eggshells with regard to Autonomy. The company aggressively defends itself against criticism from individuals who do not understand the mathematics upon which IDOL rests.

The question that many ask is, "Does IDOL work?" The answer is, "It depends." Few questioners are happy with this response. IDOL works well when several basic conditions are met.

Table 5: First, the system must be trained. The more diverse the content stream, the more effort must be put into the training process. Get the training set and tuning wrong, and IDOL will not deliver on point results when deployed. Make the investment, then the system delivers useful result sets.

Table 6: Autonomy Technology Checklist

Attribute	Verity Asserts	ArnoldIT Comment
1 Platform	Linux, Unix and variants from IBM and HP, Microsoft Windows	
2 Keyword search	Supported	Autonomy has added robust support for keyword search since 1996
3 Text mining	Yes	Autonomy generates analytics for licensees. Via the application programming interfaces, third-party analytics systems can use IDOL-generated information
4 Automated indexing	Yes	An important caveat is that the IDOL system must be properly trained and retrained to ensure that current content matches the representation models built during prior training sessions
5 Personalization	Yes	With he acquisition of Interwoven Autonomy acquired additional content management and personalization capabilities
6 Workflow	Yes	Autonomy enforces a basic content processing workflow upon licensees. Via IDOL functions and APIs, additional workflow operations can be added to IDOL
7 Interface	Graphical	Some IDOL functions require editing configuration files and writing or editing scripts. Original code may be require in some deployments
8 Hosted service	Yes	With the acquisition of Zantaz, Autonomy gained hosting innovations and capabilities
9 Administrative interface and tools	Graphical	
10 Application programming interface	Yes	IDOL offers application programming interfaces. These do not permit the licensee to interact with the core technologies in the “black box” that is IDOL
11 Professional services	Yes	With the acquisition of Verity, Autonomy gained additional consulting and engineering services resources
12 Security	Yes	IDOL supports a licensee’s security system. Additional security functions can be integrated into the IDOL system.
13 Connectors	Yes	IDOL provides connectors to major files types and enterprise content management systems such as EMC Documentum an d FileNet, among others
14 Support for structured data	Yes	Autonomy converts any content to an XML representation
15 Relevance ranking	Yes	Autonomy’s relevance is based on probabilities generate by the Bayesian-Laplace and other mathematical methods, not word frequency

Attribute	Verity Asserts	ArnoldIT Comment
16 Video	Yes	Autonomy offers video search and management systems as an extension of IDOL
17 Federated search	Yes	Queries can be passed against one or more indexes or representation of content, regardless of file type
18 Fielded search	Yes	IDOL permits field specific search
19 Content crawler	Yes	
20 Price	Begins at $500,000 but some installations reach several millions of dollars	Autonomy acquired the lower-priced Ultraseek search system with its acquisition of Verity

Second, the IDOL system requires appropriate resources. Few organizations are prepared to deliver on this condition. Computing infrastructure, bandwidth, programmers, and system administrators have to be provided. The more content that must be processed, the greater IDOL's appetite for resources.

Third, the licensee has to be prepared to check the system, perform system retraining, and ensure that the content processing pipeline is properly configured. If content transformation gets off track, IDOL will manifest certain relevancy discontinuities.

Not surprisingly, Autonomy sales professionals pursue large projects aggressively. The company's revenue growth depends on big deals, renewals, and continual stimulation from the acquisition of companies whose customers are prospects for an upsell. As a consequence, Autonomy evokes strong reactions among its supporters and detractors.

One point is clear. Autonomy was successful in generating revenue growth and staying business as competitors like Convera, Delphes, and Entopia went out of business. Autonomy had better management than companies like Endeca and Fulcrum which struggled to match Autonomy's revenue growth. Autonomy's technology proved to be more compelling than systems that were built on tried-and-true keyword methods that are known to be unsuited to certain enterprise information applications.

Finally, Autonomy recognized that Google's advertising -centric model was inappropriate for the enterprise. Autonomy focused on the enterprise sector, a path Fast Search & Transfer copies. However, Fast Search was not able to keep pace with Autonomy. As I write this in 2007, rumors of financial problem at Fast Search & Transfer may signal the demise of another search vendor unable to cope with Autonomy's business approach and technology.

Possible Drawbacks

The principal drawback of Autonomy is that licensees may not commit to the training process and then the retraining required as content drift occurs. IDOL works when the corpus is "tight"; that is, the content fed to IDOL matches the models created during training. As other types of content are fed into the system, retraining may be necessary. The IDOL system will operate in automatic mode, but tiny initial errors can, over time, have an impact on the accuracy of the metadata generated by the system. Drift is inherent in human utterance. Nuances of meaning are embedded within content, and the IDOL system benefits from on-going administrative actions. Retraining and inspection of knowledge bases generated by the IDOL system are useful activities. Unfortunately, these activities add to the cost of an already-expensive system. Is this Autonomy's fault or is it the responsibility of the licensee to perform Autonomy-recommended maintenance? The answer once again depends on one's point of view.

Not surprisingly, different licensees experience difference types problme.s. A blended corpus of general business information requires more baby sitting than a corpus composed of technical journal articles about children's heart disease. Prospectively licensees will want to conduct a thorough content inventory and then match the costs of processing that content with the IDOL system.

Other drawbacks include:

- Autonomy does not deliver a search appliance like the Google Search Appliance. IDOL must be installed, provisioned, configured, tuned, deployed, and maintained. Short cuts are not part of the IDOL methodology.
- Programmers without in-depth knowledge of IDOL may not be able to derive maximum benefits from the system. Misconfiguration can generate results that may be off point due to mistakes in training and configuring the system.
- Autonomy converts products acquired by Autonomy to IDOL. The details of specific integration are important. The prospective licensee must verify that needed functions are available within the IDOL environment so that existing applications can operate without having to be recoded to tap the IDOL version of the acquired product.

Finally, Autonomy's growth appears to be dependent on retaining existing customers and acquisitions. The approach appears to be working. However, the company is marching toward $700 million in

revenues. The key question is, “Will Autonomy be able to maintain its momentum or will it be forced to sell to a larger company?”

Anticipated Benefits

The key benefit of selecting Autonomy as a search provider is that its automatic approach to most indexing tasks can reduce deployment and tuning costs for certain use cases. Like Convera and Verity, Autonomy has a track record in search and satisfied clients like the U.S. Department of Homeland Security, Deutsche Bank, UK intelligence and police agencies.

Other benefits include:

- Autonomy, despite its financial ups and downs, has many high-profile clients. Some of these firms share their technical knowledge with other IDOL licensees. The company is unlikely to go out of business like Delphes or Entopia.
- The technology is quite different from that of most of the alternative search systems. If keyword search does not deliver, Autonomy’s Bayesian-Laplace approach warrants a closer look
- When matched to an appropriate use case and properly resourced, IDOL delivers useful outputs, particularly when other methods of information access are not able to make sense of the content objects processed by the system.

Net Net

Autonomy is a safe choice for large institutions. The company offers a system that operates automatically and largely without the costs of paying humans to read, index, and process information.

However, Autonomy is not suitable for some types of information retrieval situations. The system is based on a “black box” of technology. Like Google, Autonomy does not want licensees reverse engineering its proprietary systems and methods. For an organization wanting to fiddle with content boosting, relevance algorithms, and the internal processes of content and query processing---Autonomy is not the system to license.

For homogeneous content processing, Autonomy may be the optimal system. If large flows of content must be processed, Autonomy when properly resourced can handle almost any volume of content. Under resource the system, and it will not provide reliable real-time outputs.

A licensee who gets involved in trying to unravel the mechanics of the Autonomy math may want to approach the task with a Ph.D. in statistical

physics or a related discipline. Autonomy is a black box for a reason. The company does not choose to explain exactly how Autonomy works.

For a more open or friendly approach to information retrieval, dtSearch or Lucene may be a better choice.

Stephen E Arnold
Minor edits to a rough draft on January 23, 2014