The Power of Autonomy IDOL

Autonomy White Paper

Autonomy®

Index

The Power of Autonomy IDOL

Successfully controlling the increasingly complex and vast amount of rich information rushing into the enterprise today requires an entirely new approach—one that uses software to form an intelligent understanding of content within a contextual framework. At the core of Autonomy's scalable, modular family of product offerings is the Intelligent Data Operating Layer (IDOL).

IDOL, the foundation of the Power product family, enables organizations to index all enterprise data to optimize for maximum performance and advanced search and retrieval. IDOL can understand all types of enterprise content—whether structured, unstructured or semi-structured—including over 1,000 content types from more than 400 repositories in over 150 languages. Over 500 operations can be performed on digital content by IDOL, including hyperlinking, clustering, agents, summarization, taxonomy generation, eduction, profiling, alerting, and meaning based search and retrieval.

Autonomy's Power product offerings deliver the critical infrastructure that enables any type of organization to manage and process all of its data. While this capability is often mislabeled as simply Pan-Enterprise search, it actually encompasses a wide range of functions and services including Process Automation, Rich Media Search, Enterprise Chaining, Knowledge Management, e-Learning and Collaboration, as well as Pan-Enterprise Search.

From the early days of computers, the inability to process human-friendly, unstructured information has posed a considerable challenge. The modern IT industry was founded on the principle that, for example, if the number in Column 3 of a database goes to zero, the computer should initiate a stock order. In other words, the position of a piece of data tells the computer what to do. This process requires a tremendous amount of effort to sort and distill unstructured information into tidy rows and columns.

Using the outmoded processes just described, however, creates an increasingly impossible task as the sheer quantity of data—particularly unstructured data–continues to grow exponentially. Unstructured information, long accepted at 80 percent of the overall volume of data, is currently estimated to grow by 62% according to a leading analyst firm.[1]

Furthermore, it has been estimated that humans generated more data in 2009 than in the previous 5,000 years combined. Upping the ante, the increased use of social media, wikis, forums, blogs, Web 2.0 applications and even email, audio and video are making it harder to force all of this information into rows and columns so traditional databases can understand it.

A Meaning Based Approach to Enterprise Information

A successful enterprise IT platform must leverage the intelligence inherent within content, integrate with all enterprise systems and facilitate a data- and application-agnostic approach. Depending on access and rights requirements, information should easily flow between company divisions, geographic locations,

1 IDC, September 2010, IDC:1014

2010 Power Customer Wins

AT&T
Enterprise wide internal deployment to search for technical information on products, including audio and video content. Forms single platform across many business units.

AXA Investment Managers
3,000 seat Process Automation deployment.

Dassault Systemes
OEM license – needed a high-end embeddable search solution primarily for full text and conceptual search.

EMC
Process Automation solution covering entire document onboarding, from ingest through indexing, categorization and tagging.

Reed Business
Full spectrum search solution for initial 1m documents.

RWE
IDOL enterprise search, with connectors, for unlimited users and documents.

US Navy
Single access to all information, only one administrator needed for 400-500,000 users, ultra-high security search.

Air Liquide
Bank of Montreal
DSM
Energy Solutions
Ernst & Young
Genentech
GFI
Goodyear
Harper Collins
Hawaii Pacific Health
Liberty Mutual
Lloyds Bank
Ove Arup
Robert Bosch
Roche
RWE
Safeway
SAIC
Swedbank
University of Manchester
Vanguard
Wolters Kluwer

vendors, products, applications, operating systems and languages. The system should support search and retrieval across all repositories that contain electronically stored information (ESI), including email, IM, voice, video and text across live content and archives, data stored in centralized corporate servers, file shares, desktops and handheld devices. Once indexed, the information should be available for use by all enterprise applications to deliver the highest business value and mitigate a host of legal, regulatory and business risks.

In examining the different approaches to the challenge of unstructured information, it becomes clear that the optimal solution does not boil down to simple legacy keyword or federated searches. Instead, the most efficient and comprehensive search is one that understands the meaning of all information. With this understanding, users can leverage the maximum value of their information for the greatest competitive advantage. Through Meaning Based Computing (MBC), Autonomy far exceeds the simple search capabilities provided by most systems today. With Autonomy, organizations can facilitate a 360-degree view of their entire corpus information to drive business-critical processes.

"The biggest technology problem is both a problem and an opportunity. I would call it complexity, both of the technology itself, and of the data. My team is frankly overwhelmed with information, more than ever before. Then there is the complexity of data, which needs to be solved through intelligent data management. To achieve all this, many IT tasks that once had to be done manually are now being automated."

—Dell's CTO Peter Prince, in a recent interview with the BBC

Power Use Cases

Challenge: Unknown Unknowns

Customer: US Department of Homeland Security

User Case Summary: Threats to national security are generally unexpected and the key pieces of information surrounding them come from a variety of unrelated sources. The Department of Homeland Security (DHS) needed to ensure the implementation of an advanced knowledge sharing solution for its agents, capable of aggregating information in various formats and from multiple geographically dispersed deployments. They also needed the technology to perform advanced functions such as analyzing the data, identifying themes and issues, and where necessary delivering intelligence to agents to alert them to new developments.

With Autonomy's meaning based technology, government analysts and investigators now use natural language to describe what they are tracking or what they have heard. Because the system analyzes and understands the concepts within text, video and audio content, rather than relying on keywords, it is able to identify patterns or clusters of words that, when combined together, can signal a code for dubious activities. With this powerful technology, employees are automatically alerted to changes or developments within a wide variety of information sources. In addition, the DHS can easily build communities of experts knowledgeable about specific topics such as technology, geography, weapons and political regimes.

"The other model [to Autonomy's] using a rules-based or explicit knowledge approach is too costly because rules must be written with the assumption that things are known. In the case of Homeland Security, the investigators are looking not only for the known, but also for the unknown (new relationships) and rules cannot accommodate that."

—Autonomy Connects with Homeland Security, Dan Rasmus, Giga

Challenge: Internal and External Search

Customer: Ericsson

User Case Summary: When Ericsson realized that the cost of managing its two petabytes of enterprise information manually was practically and financially unsustainable, the company decided to build a fully scalable system to automate the management and dissemination of all data assets, including emails, media files and web pages. Ericsson decided to transform the way it managed data by creating an internal information portal and external website to help people utilize the company's wealth of knowledge.

Ericsson chose Autonomy's IDOL Server to allow all information sources to be available through one interface, providing users with an automatically personalized splash page tailored to their area of interest and expertise. On the external website, Autonomy operates transparently behind the interface to maintain their familiar look and feel, while powering search through the site's various search boxes.

As a result, IDOL increased the speed of information retrieval by an average of 58 percent, increased ranking quality by 100 percent and user satisfaction by 128 percent. Duplicate work processes were reduced by a staggering 96 percent. Overall, the company realized over $17 million in ROI in the first year of deployment.

"Autonomy technology is intricately wedded with the way users naturally work, allowing employees to identify relevant expertise within the company and boosting collaboration. Quite simply, our people spend far less time finding more relevant information."

—Nils Henström, Director of Global IT at Ericsson

Challenge: Automated Monitoring and Alerting

Customer: Precise

User Case Summary: Precise needed to monitor the entire media spectrum, including newspapers, magazines, newswires, websites, online rich media, television, radio, social media and discussion forums to provide a fully comprehensive view of media coverage and profile, and with the overall aim of developing and marketing next-generation media intelligence services to the PR and communications sectors.

Traditional search capabilities fell short of delivering the capability Precise required in order to meet client needs for content accuracy and relevance, in a time frame that supported strategic decision making.

Autonomy's core IDOL infrastructure software was selected by Precise owing to its ability to go beyond simple keyword search and provide unmatched conceptual capabilities, format and language independence and unparalleled scalability. Autonomy's technology has automated processes that were previously performed by costly and labor intensive practices, enabling Precise to enhance its media offering and liberate staff for higher value activities.

The IDOL platform, which ingests up to 500,000 news stories, programs and social media posts each day, identifies the relationships between content and client requirements on the basis of processes such as hyperlinking, clustering and categorization. Hyperlinking suggests content that is contextually linked thus ensuring clients have full visibility over related content; information clustering provides a bird's eye view of data grouping similar items together to spot new or hot trends and topics; and categorization of content automatically generates taxonomies and instantly organizes the data reducing the need for human intervention.

IDOL Technology Highlights

At the heart of Autonomy's infrastructure software is its Intelligent Data Operating Layer (IDOL) Server. The IDOL Server collects indexed data from connectors and stores it in its proprietary structure, optimized for fast processing and advanced data retrieval. As the information processing layer, IDOL forms a conceptual and contextual understanding of all content in an enterprise, automatically analyzing any piece of information from over 1,000 different content formats. IDOL offers over 500 different functions, which can be performed on any digital content, including hyperlinking, agents, summarization, taxonomy generation, clustering, eduction, profiling, alerting and retrieval.

$$H = -\Sigma p_i \cdot \log_2(p_i)$$

$$P(\theta \mid x) = \frac{P(x \mid \theta) \cdot P(\theta)}{\Sigma_{\theta' \in \Theta} P(x \mid \theta') \cdot P(\theta')}$$

Autonomy's IDOL server is built upon the seminal mathematical works of Thomas Bayes and Claude Shannon, as well as on a range of innovations that are covered by 170 patents. Autonomy technology identifies the patterns that naturally occur in text, voice or video files based on the usage and frequency of terms that correspond to specific concepts. This enables IDOL to build a precise probabilistic map of the different concepts held within a piece of data. Using this concept map, IDOL can work out the conceptual similarities between concepts contained in other pieces of data and return relevant results when a user searches for a particular term. All of this is achieved without arbitrary or subjective keywords or metatags.

Autonomy maintains an open philosophy regarding the techniques used to optimize its technology, whether they are old or new. Autonomy embraces traditional and legacy methods such as keyword, Boolean, parametric and others. However, Autonomy is best known for its pioneering work in conceptual search that is based on computational pattern recognition (non-linear adaptive digital signal processing) and contextual linguistic analysis.

Some vendors only offer "black box" solutions, believing that their technology can always provide the best answers with no tuning required. However, this idea demonstrates a naïve understanding of enterprise demands; for not even the best of automated systems can anticipate the special needs of each enterprise. These "black boxes" offer only a few, if any, tuning options for relevancy and do not reveal how the results are generated. In stark contrast, Autonomy's technology provides the best of both worlds, automatically retrieving the most accurate results using its conceptual understanding of content and also offering the flexibility to modify the relevancy algorithm, if needed. The computational process is fully transparent to the administrator and Autonomy reveals the basis for its determinations through easily understood representations such as dominant terms and idea distances.

Connectivity as a Differentiator

As has been discussed, there is a constant growth in both the quantity of information and the number of source formats. The first step to leveraging this resource is being able to connect to and access the different types of information. This connectivity is far easier said than done, and it is true that most vendors claim a number of connectors or SDK's to develop more, but this is not the same as truly connecting to all data types, maintaining existing enterprise security models and allowing support for legacy file systems and future emerging standards.

Autonomy uniquely owns its own technology for accessing 400+ file repositories and 1,000+ individual file formats and is the clear and acknowledged OEM leader in this area, providing technology to over 400 partners to integrate into their own solutions.

Process Automation

By using a combination of Autonomy's filtering technology and intelligent IDOL Server, organizations can automate previously time-consuming tasks and processes. For example, the processes of on-boarding and product cataloguing are often labor intensive tasks. They require humans to read product descriptions, extract particular information and descriptions and then apply

appropriate tags to ensure the product appears correctly in a cataloguing system. With Autonomy, these types of procedures can be automated by applying various functions such as OCR (on both electronic data and hard copies), tagging, categorization and normalization at the first point of contact with new product information. Once content has been indexed by IDOL, additional functions can be applied such as hyperlinking, clustering and or alerting to connect relevant employees in the organization.

Through Autonomy IDOL, this automation at both the input and output points of file processing can be embedded within existing applications or alongside existing API's to directly speed organizational processes without requiring extensive software platform migration or staff re-training.

Autonomy's connectors uniquely connect to over 400 different content repositories and, by using API integrations and mimicking client connections, full metadata and full text is extracted along with native security credentials for each repository. This allows Autonomy to extract full content from each repository instead of relying on the inferior approach used by many vendors that involves using web spiders to gather content. In addition, this enables massive scalability, along with the benefit of mapped security, as security fields are fully mapped upon ingestion.

Enterprise Integration

Because all of Autonomy's technology, in all three product areas—Power, Protect and Promote—is built on the same IDOL platform, Autonomy can offer unrivalled opportunities for enterprise integration on a single platform. This enables an organization's entire IT installation to share a single platform allowing all modules and systems to communicate seamlessly across the enterprise. The importance of this approach is proven by the hundreds of companies that have chosen to standardize on Autonomy for not just their general Power infrastructure, but also for their Protect and Promote implementations.

As a result of IDOL's flexibility and small footprint, it is also easy to embed the technology behind the scenes in website search boxes, portlets, API's and other applications. This allows users to maintain the look and feel of a familiar interface, while making use of a single access point to the entire corpus of data with availability for IDOL's intelligent search functions.

Unstructured Information

Since Autonomy has the most complete set of connectors on the market, allowing it to access information in all forms, and because it forms a single, seamless platform across an organization's entire information assets, IDOL can process and leverage structured, unstructured and semi-structured information, no matter where it resides. This is particularly important with the increased popularity of rich media including images, video and audio; all of which IDOL can categorize, index, retrieve and forward in exactly the same way as text. By relying on probabilistic nearness, IDOL also avoids the use of tags or metadata, which are often incomplete and always subjective. It is also important to note that most lower-end solutions which claim to offer conceptual search in fact rely on these same inconsistent tags to generate results.

2010 Technology Updates

Results Relevancy and Defensibility

- Fully transparent computational process
- Weight of every field can be manipulated
- User feedback used to calculate relevancy and primacy
- Recently added features include:
  - Query bread-crumbing/Audit
  - Real-time validation sampling
  - Defensibility reporting
  - Intent-based ranking
  - Social media scoring

New Enhancements to IDOL Connector Framework

- New and improved connectors to over 400 content repositories, including EMC Documentum, IBM WebSphere, Microsoft SharePoint, Lotus Notes, Oracle BEA WebLogic and file systems
- Normalizes metadata to enable unified actions against all content repositories
- Provides enhanced audit capabilities and document tracking to automate the legal hold and collections process, reducing the risk of human manual error
- Extends Autonomy's unique "Manage-in-Place" technology to a much broader set of content repositories
- New capabilities for automatically publishing content to customer-facing systems, including websites, online advertising campaigns, email systems, and customer relationship management applications
- Automated self-scaling and load balancing capabilities for reduced IT overhead and improved performance

Architectural Improvements

- Distributed link architecture
- Intelligent repository storage
- Distributed hot deployment mode
- Distributed abridged mode

Language

In addition to accessing and processing almost any type of content, IDOL can connect to content in 150 languages, including double-byte languages. In the same way that IDOL builds a mathematical model of rich media content without relying on specific words, the actual language of words is not important. If further language models need to be added, this is a trivial matter of matching set linguistic data to the indexed content.

Language agnosticism makes it possible for a user to search in any language, enter a search term in one language and return results in another language and even provide a conceptual summary of content in other languages. This capability has been used for multi-channel media monitoring where mentions of specific terms must be identified, regardless of the language.

Accuracy and Learning Ability

Building on the flexibility of IDOL to easily build in a new language model, it is also very easy to add in existing corporate thesauri, style guides and taxonomies. As a result, an organization can personalize its IDOL installation and ensure that all natural language searches return the most accurate and relevant results, without ignoring years of developed corporate language. While conceptual search is a powerful tool, it cannot replace human expertise and knowledge built up over years. When these two strengths are combined, the results are even more impressive than when either set of knowledge is applied alone.

In addition to simply learning information that is taught, as with taxonomies or language models, IDOL can learn by adapting to content as it is authored and used. For example, over time the meaning of words and phrases can change, such as "ground zero," which originally meant the fallout zone from a nuclear weapon, but has since adapted to mean the site of the twin towers after 9/11.

Security and Scalability

As discussed under Automation, Autonomy can uniquely provide highly secure mapped security by embedding IDOL within existing applications and alongside API's to extract full, native security credentials for the specific content and repository. As a result, IDOL eliminates the need to send any requests across the network to data stores when building up a search results list which ensures no degradation of performance. Access rights are assessed 'in line' within the IDOL kernel at speeds that exceed the response time of the native repository. IDOL uses 128-bit encryption and can leverage SSL for aggregation and querying of content. Autonomy's proven technology has consistently demonstrated its ability to deliver highly secure search for customers including the US Department of Homeland Security, the US Department of Defense, the UK MoD and other government and intelligence agencies worldwide.

As previously stated, the amount of unstructured information is growing exponentially, requiring organizations to choose enterprise infrastructure solutions that are highly scalable to cope with this increasing demand. A single IDOL engine is capable of supporting over 250 million documents on 64-bit platforms, indexing over 60 gigabytes per hour with guaranteed index commit times of less than five milliseconds. A single IDOL engine can execute 2,000 queries per second, across the entire corpus of indexed data, with sub-second response times.

Autonomy is trusted to power some of the world's largest IT deployments, including:

- ***US DoD****– Over 2.5 million users*
- ***Bloomberg****– 276,000 users and 46 million emails per day*
- ***ChoicePoint****– 4.8 billion documents indexed and stored in Autonomy cloud-based data centers*

Autonomy operates the world's largest private cloud with currently over 17 petabytes of data under management, growing by three million files per hour and operating across 7,000 production servers hosted in seven state of the art data centers around the globe.

Industry Validation for IDOL

Ovum Decision Matrix for Enterprise Search and Retrieval (2010)

In Ovum's 2010 Decision Matrix for Enterprise Search and Retrieval, **Autonomy was the only vendor to be awarded the highest ranking: "a company to be shortlisted."** The report evaluates 12 vendors including Endeca, Google, IBM, Microsoft, Oracle, Recommind, Exalead, Vivisimo, Sinequa and identifies Autonomy as the clear leader in the ESR market.

"Autonomy is the clear leader of the ESR market and is an automatic shortlist into any vendor selection exercise...it is a natural fit for organizations that are looking at ESR from a strategic perspective."

"The vendor is an order of magnitude larger than most other peers in ESR and delivers the most sophisticated and comprehensive search functionalities among peers."

"The vendor edges past competition by offering the most granular of functionality natively in its suite, while others offer the same through integration with third-party products."

"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 solution's functionality is deemed to be the best-in-class."

"Ovum is particularly impressed with the level of automation IDOL enables, both within a search platform and across the organization."

IDC Worldwide Search and Discovery Software 2008-2012

IDC's "Worldwide Search and Discovery Software, 2008–2012 Forecast Update and 2008 Vendor Shares Bloom Amid Economic Gloom" evaluates and ranks the top vendors in enterprise search.

IDC ranked Autonomy as the clear market share leader (15.5%), holding a commanding market share that is nearly double that of Google (8.5%) and nearly five times that of Microsoft (3.6%). This extensive leadership position comes despite the report lumping together all types of search solutions from the low-end market. Autonomy easily has over 90 percent of the market share in the target market of high-end, global enterprise solutions. In addition, the worldwide search and discovery market is expected to continue to grow at a faster pace than the overall worldwide software market, at 12.9 percent versus 4.4 percent in 2009.

"When it first conceived IDOL platform as a kind of "information bus," well over six years ago, the logic and breadth of this idea was too early for the market. Autonomy uses its IDOL platform to process unstructured information of all types. It sees search as a foundation technology, such as databases, that, combined with workflow, knowledge bases, rich media technologies, or collaboration can give a vendor the flexibility to produce specific applications that address departmental or cross enterprise problems. Its product suite is strong in rich media, from its eTalk and Virage applications, and in search, pattern matching, workflow (Cardiff) compliance, and email archiving. At this point, it is clear that Autonomy should no longer be considered purely a search vendor. It builds search-based applications to answer market demands for better information-centric software."

Independent Broker Research Report, Bryan Garnier & Co (2011)

Power: Automating without Re-Hiring

The 2007-09 economic crisis was an excellent catalyst for Autonomy, as it was a way for customers to make significant productivity gains with fewer people. Slowly emerging from the crisis, customers continued to automate Enterprise Search as a means of processing much more information without hiring back all the staff they had before. On the corporate side, given the higher level of maturity of the market, Autonomy estimates it may generate 10-15% growth when the economy is healthy.

According to Gartner, the relevant market (Information Access Technology) amounted to US$1.1bn and is expected to increase by an average of 12% per year until 2013[2]. However, this segment can be classified as discretionary spend as it is dependent on the number of users, so in a downturn growth is likely to slow. That said, we consider Autonomy still has significant market penetration potential ahead, as only 100 customers have standardized on the IDOL platform. Once Autonomy has "laid the foundations" in the customer's core IT platform, there is room to grow.

Autonomy leads the pack in the Power segment, as IDOL is probably the most complete and high-end product offering (documents, e-mails, audio, video…). Followers in the high-end segment are mainly Endeca and Dassault Systèmes (Exalead), **but in 90% of the deals Autonomy has no real competition.** Google (with Search Appliance), Microsoft (since the acquisition of FAST in 2008), IBM and Oracle are also competitors, but either they target mainly small to mid-sized deployments (Microsoft is focused on its SharePoint web portal software) or their scope is limited.

"Microsoft, Oracle, IBM, and Endeca form a distant second rung to [Autonomy]."

—OVUM, Aug 2010

"Autonomy dominates search space, reports IDC."

—Computer Weekly

"Autonomy won the enterprise search wars."

—Computer Business Review

"Autonomy is the market leader in the provision of software that automates the analysis of unstructured data, whether in the form of text, audio, images or video."

—UBS

"Autonomy is the market leader in providing software to automate the classification and management of data."

—Margaret Lawson, SVM, Apr 2010

"We see Autonomy as a structural growth story. The company with its IDOL platform is one of a few vendors that can help corporations to make sense of an ever-increasing amount of unstructured data."

—Barclays Capital, Oct 2010

2 Gartner, August 9, 2010 ID:G00205432

"Autonomy Process Automation allows our company to operate with the speed and agility we need to provide superior service to our clients."

—Scott Hanson, Baird, Jul 2010

"The need to digest and assess information swiftly and accurately so the correct decisions can be taken as quickly as possible has never been greater... and Autonomy's unique search engine capabilities are providing their worth in this field."

—Russ Mould & Jessica Fourseth, Shares, June 2010

" By enabling us to move many of these paper-intensive processes into a secure online environment Autonomy Process Automation has significantly reduced the time and costs associated with the management of our business processes."

—Kara Rubinson, Palm Beach County School District, Jun 2010

About Autonomy

Autonomy Corporation plc (LSE: AU. or AU.L), a global leader in infrastructure software for the enterprise, spearheads the Meaning Based Computing movement. IDC recently recognized Autonomy as having the largest market share and fastest growth in the worldwide search and discovery market. Autonomy's technology allows computers to harness the full richness of human information, forming a conceptual and contextual understanding of any piece of electronic data, including unstructured information, such as text, email, web pages, voice, or video. Autonomy's software powers the full spectrum of mission-critical enterprise applications including pan-enterprise search, customer interaction solutions, information governance, end-to-end eDiscovery, records management, archiving, business process management, web content management, web optimization, rich media management and video and audio analysis.

Autonomy's customer base is comprised of more than 20,000 global companies, law firms and federal agencies including: AOL, BAE Systems, BBC, Bloomberg, Boeing, Citigroup, Coca Cola, Daimler AG, Deutsche Bank, DLA Piper, Ericsson, FedEx, Ford, GlaxoSmithKline, Lloyds TSB, NASA, Nestlé, the New York Stock Exchange, Reuters, Shell, Tesco, T-Mobile, the U.S. Department of Energy, the U.S. Department of Homeland Security and the U.S. Securities and Exchange Commission. More than 400 companies OEM Autonomy technology, including Symantec, Citrix, HP, Novell, Oracle, Sybase and TIBCO.

Please visit **www.autonomy.com** to find out more.

Autonomy Inc. and Autonomy Systems Limited are both subsidiaries of Autonomy Corporation plc.

20110214_PI_WP_Power_IDOL

Autonomy