Autonomy IDOL 10

IDOL10™

Understanding Structured and Unstructured Data in Real Time

The information powering today's organizations exists in two forms: structured and unstructured. When computers were first used, the focus was on structured data, which required humans to adapt the information to fit machines. Manual classification augmented structured information with the nuances and complexity that computers could not grasp, because people do not speak in zeroes and ones. The nature of human communication is complex, using language and idioms, photographs and videos, recordings and social media interactions. Human information is unstructured and does not fit into the neat rows and columns of relational databases. Yet, it represents the fastest growing segment of the world's content.

Now, a groundbreaking information platform, Autonomy IDOL 10, enables organizations to understand and process 100 percent of enterprise information, structured and unstructured, in real time.

Architecture of Structured Information and the Unstructured Data Enterprise

Addressing the need left unanswered by traditional databases to understand human information, Autonomy has leveraged Meaning Based Computing to create its next-generation platform built for the human information era. Autonomy takes computing to the next level by understanding and extracting meaning and insight from any type of data—most notably the unstructured portion that makes up approximately 85 percent of content today.

In the modern enterprise, information is captured, monitored, and created in myriad forms: in databases, from websites, and in emails, audio, video, social networking, blogs, call-center conversations, CCTV footage, and sensor data. Without technology that can automatically understand this information, a tremendous amount of manual effort is required to sort, distill, and tag unstructured content, or organize and migrate structured data within legacy databases or data warehouses. Now, however, organizations are realizing that structuring information using excessive manpower and costly infrastructures is not the answer. The answer to leveraging information to its highest value requires technology that understands information in all its forms, and in real time.

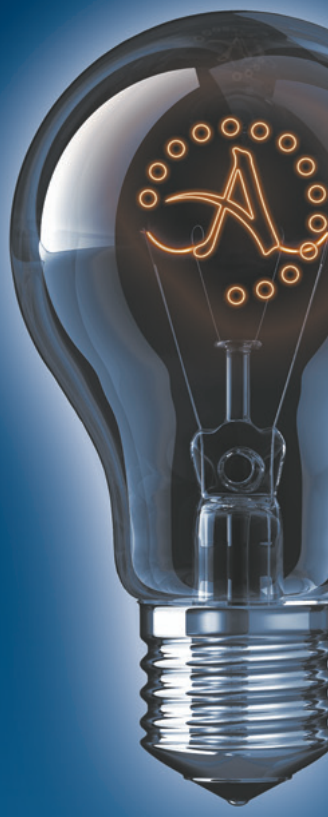
High-Speed, Real-Time Analytics

Today, the combination of Vertica's high-speed analytics platform with Autonomy's IDOL technology marks a fundamental shift in our ability to process the volume and diversity of data in organizations today. We are at an historic moment when it is the "I" in 'Information Technology' that is changing. Autonomy provides solutions that understand the full spectrum of enterprise information, both human and structured, as well as the relationships that exist within it. By enabling computers to understand the shades of grey in our world, rather than only the black and white found in databases, Autonomy Information Management allows businesses to automate key processes and improve efficiency.

"Autonomy IDOL 10 brings these worlds together, allowing organizations to automatically process, understand, and act on 100 percent of their data, in real time. The results will be dramatic, as businesses can develop entirely new applications that explore the richness and color of Human Information that live in unstructured, semi-structured, and structured forms."

Yvonne Genovese, Gartner

Autonomy®
an HP company



POWER
PROTECT
PROMOTE

The Next-Generation Information Platform Built for the Human Information Era

Autonomy's next-generation information platform, IDOL 10, is a single processing layer that enables organizations to extract meaning and act on all forms of information, including audio, video, social media, email and web content, as well as structured data such as customer transaction logs and machine-based sensor data. The platform combines Autonomy's infrastructure software for automatically processing and understanding unstructured data with the high-performance, real-time analytics engine for extreme structured data from Vertica, an HP Company.

The Platform Built for the Human Information Era: IDOL 10

Autonomy IDOL 10 offers a range of advantages to help today's modern enterprise leverage every form of information to its highest value.

- *A single processing layer for forming a conceptual, contextual and real-time understanding of all types of data, both inside and outside the enterprise.*
- *A combination of Autonomy's infrastructure software for automatically processing and understanding unstructured data with Vertica's high-performance real-time analytics engine for extreme structured data.*
- *Unique pattern-matching technologies, powered by an analytics engine based on statistical algorithms that recognize distance in ideas as well as concepts and context in real time.*
- *Five new solution sets – HP Big Data Solutions, HP Social Media Solutions, HP Risk Management Solutions, HP Cloud Solutions, and HP Mobility Solutions.*
- *"Manage-in-place" technology, which forms an index of all forms of data, allowing information to reside in its original location. This eliminates the need to make copies of data, reduces storage and hardware costs, and mitigates the risk involved in performing inefficient transfers of data.*
- *A NoSQL interface provides a single processing layer to perform cross-channel analytics on both structured and unstructured data.*
- *The Vertica Analytics Platform, which includes enhanced native in-database analytics, provides new capabilities for geospatial, event-series pattern matching, event-series joins, and advanced aggregate statistical and regression analytics.*
- *Vertica's real-time analytics for real world applications delivers performance enhancements throughout the Vertica Analytics Platform in areas such as sub-queries, database statistics, lifecycle management, query optimization, data re-segmentation and join filtering.*
- *Enhanced elasticity features enable dynamic expansion and contraction of clusters more than 20 times faster in every deployment scenario – cloud, virtual and physical – allowing users to quickly create additional capacity as needed.*

Key Idol Functions

IDOL offers has over 500 functions for processing information. The IDOL functions listed here are some of the most powerful and widely deployed.

Automatic Hyperlinking – IDOL allows manual and automatic linking of related pieces of information, regardless of format. Concepts in documents can be linked automatically to those in another file, or to related concepts within video or voice mail. Hyperlinks are generated in real-time at the moment a document is viewed, eliminating the need for manual intervention and ensuring hyperlinks are up-to-date.

Automatic Categorization – IDOL categorizes concepts found within unstructured text, ensuring that all data is classified in the correct context with the utmost accuracy. Existing legacy taxonomies can be maintained or enriched with contextual understanding. Categorized information is presented to users in the form of channels.

Automatic Query Guidance (AQG) – AQG automatically provides keyword-oriented users with the suggestions for finding the most relevant information, and identifies the different meanings of a term by dynamically clustering the results into their most relevant groupings. Users can guide the engine simply by clicking on the context that most applies to their interests.

Automatic Taxonomy Generation – IDOL forms an understanding of enterprise information, automatically generates taxonomies, and instantly organizes data into a familiar child/parent taxonomical structure.

Real-Time Sentiment Analysis – IDOL analyzes the structures and meaning of language to determine the positive and negative characteristics of each piece of information and creates relevant classification systems. IDOL can determine the degree to which a sentiment is positive, negative, or neutral for the entire content or a segment of the content. Administrators can apply multiple tagging functions and specific threshold cut-offs to determine the sensitivity of sentiment analysis.

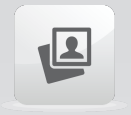
Automatic Clustering – Organizations can analyze large sets of documents and even user profiles and automatically identify inherent themes or information clusters. IDOL even clusters the unstructured content exchanged in email, telephone conversations and instant messages. IDOL uses the most advanced heuristics, such as quantum clustering, to form these conceptual groupings.

Directed Navigation – IDOL dynamically presents navigable parameters based on document properties and all metadata attributes: explicitly defined, automatically generated, and even fields not previously used. Directed Navigation offers organizations a way to integrate valuable information from structured databases with unstructured content.

Eduction – Beyond traditional entity extraction, IDOL enriches extracted data based on knowledge in the organization. IDOL can extract specific data such as organizations, people, places and figures, and conceptually relate the information to other data in the enterprise and automatically form relationships. Eduction includes out-of-the-box entities such as names of commercial organizations, people, places, postal/Internet addresses, phone numbers, dates, times, numbers, prices, Social Security numbers, job titles and holidays.

Personalized Agents – Personalized Agents monitor information for users 24/7 on specific topics from a range of data sources, such as persistent interests. Agents can be defined or trained explicitly with a natural language description or Boolean expression. Agents can be trained or re-trained by example, simply by being shown a document, video, or verbal conversation that matches a user's interests, and can learn the concepts within the example and define itself accordingly. Once created, Agents monitor changing information in real-time within the index, instantly alerting the user to anything new that matches the training.

Profiling – IDOL accurately understands individuals' interests based on their browsing, content consumption, and content contribution. IDOL generates a multifaceted conceptual profile of each user based on both explicit (agents) and implicit profiles (click-through and submission), and forms a very current understanding of users' interests and personalizes the relevancy model to deliver intent-based search results.



Social Search – IDOL provides a rich participatory platform for capturing unstructured, tacit knowledge created from Web 2.0 channels, connecting people to related communities, projects, and processes. Users can contribute and share tags, comments, and votes, and share search folders and knowledge with colleagues. User activity, explicit and implicit, can be incorporated to influence the relevancy calculation.

Conceptual Retrieval – Built on an innovative pattern-recognition technology, IDOL offers higher degrees of accuracy and sophistication using scalable technology that recognizes concepts, and does not rely on simply words in the document.

Full XML Support – IDOL eliminates the manual inefficiencies of XML tags by understanding the content and purpose of the tag itself, related information, or both. IDOL automatically inserts XML tags and links into documents based on concepts contained in the information, and uses meaning-based technology to provide an infrastructure for complete and automatic interoperability between applications using different XML tagging rules.

Autonomy Business Console (ABC) – ABC empowers business users to react quickly to changing usage patterns and provides highly targeted results to end-users without programming. ABC uses a robust set of widgets, wizards and tools to simplify the management of IDOL functions, including promotions, synonym management and relevance boosting.

IDOL Echo – IDOL Echo allows fully auditable and accountable monitoring of information use, enabling the enterprise to forensically account, track, and trace each piece of data that enters, leaves, is born, or dies within the organization. Echo follows a file's path and history and can report on who or what the asset has influenced. It can not only follow a traffic pattern such as the path of an email attachment or voice mail (i.e. what, when, who read, heard, forwarded, retained), but also detect the influence of content (i.e. who has taken, re-purposed, or even plagiarized).

IDOL Architecture

IDOL provides the architecture requirements of a pan-enterprise search platform, including the following:

FRCP-Compliant Search – The Federal Rules of Civil Procedure (FRCP) render all relevant Electronically Stored Information (ESI) discoverable, regardless of format or location. To be FRCP-compliant, a pan-enterprise search platform must:

- ***Search ALL repositories***
- ***Search the entire repository without relying on "jump-out"*** *– This performance-enhancing tactic used by some vendors stops searching an index as soon as it assembles a group of results that has been deemed large enough, without searching the entire index*
- ***Produce auditable results*** *– All data must be searched fully to be considered auditable*
- ***Pass authentic results to a hold function*** *– Relevant ESI cannot be altered in any way or deleted*

Scalability and Performance – IDOL scales to support the largest enterprise-wide and portal deployments in the world, with presence in virtually every vertical market. Since IDOL's scalability is based on its modular architecture, it can handle massive amounts of data on commodity dual-CPU servers. IDOL delivers linear scalability through a multi-threaded, multi-instance approach with load-balancing to intelligently distribute the indexing and query workload.

Mapped Security – The biggest single constraint on scalability within enterprise applications is the ability to manage entitlement checks in a scalable manner. IDOL stores security information in its native form directly in the kernel of the engine itself, with automatic updates to keep the security data current. This sharply contrasts with other security models that store security information in the original repositories, requiring communication between the search engine and the underlying repository for every potential result at the time of query.

Global Language Support – The need to manage content in varied languages has never been more acute for today's global enterprise. IDOL's language-independent technology develops a statistical understanding of the patterns of any language using sophisticated probabilistic modeling and pattern-matching techniques. It currently supports over 150 languages and provides cross-lingual search. Autonomy also provides optional language packs to further enhance localization, including stemming, stop lists, transliteration, multiple encoding support and term decomposition.

HP Information Optimization is a core component of an Instant-On Enterprise. In a world of continuous connectivity, the Instant-On Enterprise embeds technology in everything it does to serve customers, employees, partners and citizens with whatever they need, instantly.

About Autonomy

Autonomy Corporation, an HP Company, is a global leader in software that processes human information, or unstructured data, including social media, email, video, audio, text, web pages and more, enabling companies to leverage their data assets.

About HP

HP creates new possibilities for technology to have a meaningful impact on people, businesses, governments and society. The world's largest technology company, HP brings together a portfolio that spans printing, personal computing, software, services and IT infrastructure at the convergence of the cloud and connectivity, creating seamless, secure, context-aware experiences for a connected world. More information about HP (NYSE: HPQ) is available at http://www.hp.com

Autonomy Inc., an HP Company
One Market, Spear Tower, 19th Floor,
San Francisco, CA 94105, USA
Tel: +1 415 243 9955
Fax: +1 415 243 9984
Email: info@us.autonomy.com

Autonomy Systems Ltd., an HP Company
Cambridge Business Park,
Cowley Rd, Cambridge CB4 0WZ, UK
Tel: +44 (0) 1223 448 000
Fax: +44 (0) 1223 448 001
Email: autonomy@autonomy.com

Other Offices
Autonomy has additional offices in Antwerp, Barcelona, Beijing, Boston, Calgary, Cambridge, Chicago, Dallas, Kuala Lumpur, London, Madrid, Mexico City, Milan, Munich, New York, Paris, Pleasanton, Rome, San Francisco, Santa Clara, Shanghai, Singapore, Santiago, Sao Paulo, Stockholm, Sydney, Tokyo, Utrecht and Washington, D.C.

www.autonomy.com

Autonomy®
an HP company