

Brochure

Next-generation information analytics

HP IDOL 10: built for the era of big data



Get your greatest return on information

We are experiencing transformational changes in the computing arena. Data is doubling every 12 to 18 months, accelerating the pace of innovation and time-to-value. The vast majority of today's data growth is driven by the creation of human-friendly information such as documents, social content, video, audio, and images. In fact, 90% of digital content being created will include these unstructured data types by 2015¹. Data growth is also accelerated by the ubiquity of mobile devices, which gives us “always on” applications and users who demand anytime, anywhere access to information.

Highlights

- **Understand virtually all of your information with high-performance analytics:** Over 500 analytical functions available for text, audio, video, and image
- **Derive actionable insights:** Process data in near real time to gain a competitive edge
- **Maximize your information reach:** Connect to over 400 systems with support for over 1000 file formats, so you can find all relevant information
- **Let social media work for you:** Detect emerging trends and influencers in this powerful media with sophisticated sentiment analysis and clustering technology
- **Achieve big data scalability and security:** Track record of supporting the largest enterprises in volume of information, number of users, and frequency of transactions while adhering to strict security standards
- **Easy to use and administer:** Leverage a visual dashboard and simplified processes
- **Manage data in place:** Manage data within the original repository while ensuring its accessibility
- **Rapid innovation:** Roadmap includes additional next-gen capabilities via quarterly releases

As a business leader, you depend on human information to keep your business working to its full potential. The information analytics platform, HP IDOL 10 (Intelligent Data Operating Layer), from HP Autonomy, has been built to address these modern challenges.

IDOL 10 offers you a single processing layer that helps you automatically unlock key ideas and concepts in all your information, structured and unstructured. Based on advanced mathematical principles and augmented by over 170 patents, IDOL 10 understand the key concepts present in virtually all forms of content, helping you to detect patterns and relationships between data.

With IDOL 10, you can understand and act upon documents, emails, video, chat, phone calls, social media, and application data at the same time and faster than ever before. Because we know information is fragmented across a variety of repositories, IDOL 10 streamlines information processing across networks, the web, and the cloud. We help you see your information wherever it is. After all, why should you work for your information when it can work for you?



IDOL 10 understands virtually all information fragmented across different silos, and finds patterns and relationships surfaced through different applications

IDOL 10 powers many of the applications that depend upon an understanding of human information to perform, including:

- Enterprise Search
- eDiscovery
- Voice of the Customer/Workforce
- Marketing Optimization
- Enterprise Content Management
- Social Media Analytics

¹ IDC Predictions 2012: Competing for 2020

“HP Autonomy, with the power of its IDOL engine, takes fan data, collects it, and stitches it together...that helps us understand what is being talked about across the ecosystem of the sport.”

—Steve Worling, Senior Director of IT, NASCAR

“Recruiters are highly skilled individuals, and that’s not something you can replace—but what we can do is give them a tool that helps them work more efficiently. With IDOL, instead of 50 loosely-related candidates to review, recruiters get the 15 most suitable, and work from there.”

—Daniel Richardson, CTO, Bond International Software

If you can't find your data, does it really exist?

We have realized that traditional ways of understanding and using information simply fall short in providing the business benefits that are required in today's world. When detectives try to uncover a crime, they look for incriminating emails and wade through surveillance videos. When marketers want to know more about their customers, they look to Twitter. Information today is diverse, dynamic, and complicated. The old method of finding answers or uncovering predictive patterns in the rows and columns of a database take far too much time, leave out gaps of knowledge, and often yield incomplete results. Similarly, keyword and metadata-based search return too many irrelevant and incomplete results, without giving actionable insights to make the information count.

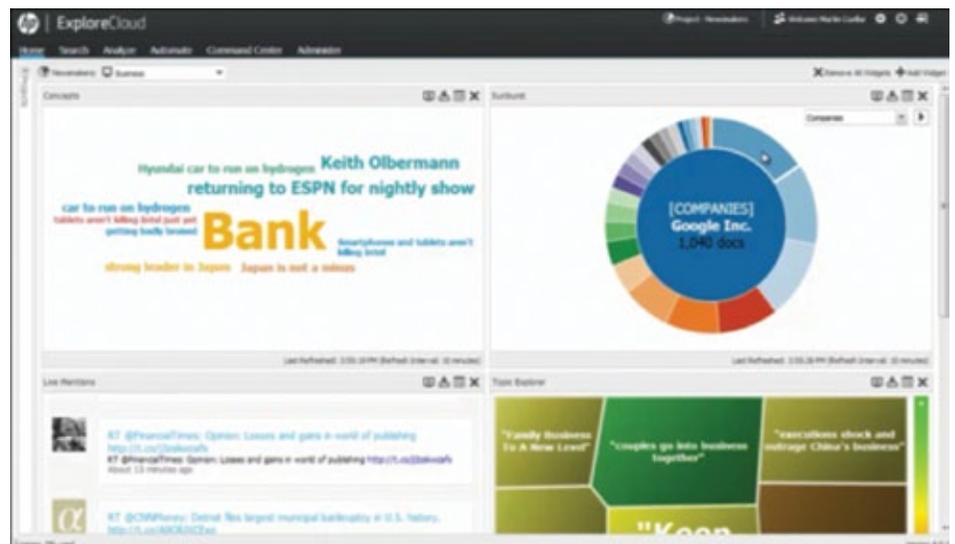
The explosion in the variety of file formats leaves most search and analytics technologies incapable of making sense of new file types. Since they rely on user-defined tags, which are often inconsistent and too vague to provide much meaning, you inevitably can't connect with information you probably have, but for all practical purposes don't know exists.

There is simply too much data to sift through all at once without using a system that can actually understand it, and this unfortunately renders a lot of great information unsearchable and unusable.

Understanding meaning helps you unleash the value of your information

When you can analyze all types of information at once, you can make decisions while there is still time to change your course. For instance, what if you needed to simultaneously analyze all customers in one zip code who use an iPhone 4, provided product feedback when they called customer service, and posted a Yelp review in the same month? If you processed the data manually, it would take so long that the information would lose its value before you could use it. With IDOL 10, you can now perform sophisticated analytics on all types of information at the same time, and faster than ever before.

IDOL 10 uses patented probabilistic and pattern-matching algorithms to automatically recognize concepts and ideas in all your information. Because the technology is based on the universal language of mathematics, it treats words as “symbols” and works in all languages. IDOL 10's unique technology detects patterns, emotions, sentiments, intentions, and preferences as they happen. Knowing more about your data in less time speeds the pace of business and strengthens your ability to compete.



IDOL 10 provides visualization tools to reveal insights about data

File Types

Search over 1000 file types, including audio and video files

Content Connectors

Over 400, including:

- Alfresco
- Confluence
- Documentum
- Ektron
- Enterprise Vault
- eRoom
- Exchange
- FileNet P8
- FileSystem
- FTP
- Hadoop
- IBM Seedlist
- IMAP
- KVS
- LiveLink
- Lync
- Notes
- Objective WebTalk
- OCS
- ODBC
- Oracle
- Oracle UCM
- PLC
- POP3
- ProjectWise
- RSS
- Salesforce
- SharePoint 2007
- SharePoint 2010
- Social Media and websites (e.g. Facebook, Google Docs, LinkedIn, Twitter)
- Sybase
- Vignette
- WebDAV
- Weibo

Get a 360-degree view with exploratory analytics

We understand that the information you find on Twitter and Facebook is becoming just as important as the information you find in contracts and spreadsheets. And to help you understand all content inside and outside the enterprise, we built a platform that gives you a 360-degree view into all data sources that are important to your business. IDOL 10 helps you stay informed, compare and control document versions, create reports, move data, and manage important information ranging from social media posts to files created in productivity tools. Having a seamless view across all your information helps you manage better, collaborate effectively, and act more quickly.

“HP Autonomy’s exceptional sophistication and extensibility enable it to tackle the most demanding use cases.”

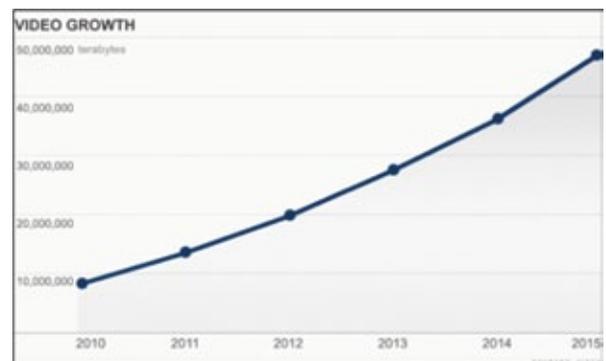
—Gartner Group

Understand rich media as easily as text

Multimedia files such as video, audio and images are proliferating at an unprecedented rate. Partly driven by the growth of smartphones and the increasing presence of CCTVs, rich media represents a significant source of insight, if leveraged beyond metatags.

Rich media as intelligence

You may think that rich media technology is confined to law enforcement agencies who must survey massive amount of video feeds, or specialty media agencies. When you consider the bombing at the Boston marathon, you can see the vital importance of video assets to this sector, and open source intelligence. When you can leverage videos captured from people’s phones and shared via social media, for instance, everyone becomes a potential contributor. In this way, rich media technology plays a critical role in intelligence.



eConsultancy 2013

Rich media in the enterprise and social networks

In the enterprise, the role of video content has grown dramatically. Video and audio use has catapulted in recent years with the abundance of meetings conducted online. There is also a massive amount of valuable video information captured in multimedia training, video conferencing, webcasts, demos, and Human Resources material.

Video has now become the primary channel for people to consume content online. For instance, people upload ranting videos on YouTube and post them on Twitter. Traditional news broadcasts may mention activities around your competitor. And of course, there are call centers whose service performance informs much of a customer’s impression of your brand.

With the rise of smartphones, countless photos are being uploaded to many social media sites. Information contained in these images reveals much more valuable information about relationships than what how you can define through social networks. For instance, who do you really spend most of your time with? Doing which activities? In what seasons? Wearing what types of clothes?

Languages

Search over 150 languages, including:

- Acehnese
- Afrikaans
- Albanian
- Amharic
- Arabic
- Armenian
- Azeri
- Basque
- Belarussian
- Bengali
- Berber
- Bihari languages (Magahi, Bhojpuri, Maithili)
- Bikol
- Bishnupriya
- Bosnian
- Breton
- Bulgarian
- Burmese
- Catalan
- Cebuano
- Cherokee
- Chinese
- Chuvash
- Croatian
- Czech
- Danish
- Divehi
- Dutch
- English
- Erzya
- Esperanto
- Estonian
- Ethiopic languages
- Faroese
- Farsi/Persian
- Finnish
- Flemish
- French
- Frisian
- Galician
- Georgian
- German
- Gilaki
- Greek
- Greenlandic
- Guarani
- Gujarati
- Haitian
- Hausa
- Hawaiian
- Hebrew
- Hindi
- Hungarian
- Icelandic
- Ido
- Igbo
- Ilokano
- Indonesian
- Irish Gaelic
- Italian
- Japanese
- Javanese
- Kalmyk
- Kannada
- Kapampangan
- Kazakh
- Khmer
- Kikongo
- Kinyarwanda
- Kirundi
- Komi
- Korean
- Kyrgyz
- Lao
- Lappish
- Latin
- Latvian
- Lingala
- Lithuanian
- Luxembourgish
- Macedonian
- Malagasy
- Malay
- Malayalam
- Maltese
- Manipuri
- Maori
- Marathi
- Mazandarani
- Mirandese

With IDOL 10, you can leverage this wealth of rich media, transforming it into intelligent assets. The full range of IDOL 10 functions can be performed on rich media and cross-referenced with any other form of information such as text. You can automatically generate metadata, keyframes, transcriptions, and understand audio and video streams in near real time. IDOL 10 allows you to delve into the video file itself—rather than relying on tags or metadata—to extract the meaning of the information contained within. It can detect objects, identify faces and logos, and perform OCR on images and even extract key fields from scanned documents.



IDOL 10 automates video and audio analysis and identifies items such as logos, detects faces, performs OCR, generates keyframes, and understands concepts

Easy to consume and administer

With IDOL 10, administrators can quickly troubleshoot performance-related issues with an intuitive, visual dashboard that allows you to remain productive. The user interface displays how resources are being used, allows query speed analysis to help identify why a query may be slow (including a graphical breakdown of where the engine is spending time), shows the status and progress of the index, highlights errors and warnings in logs and in the configuration file, and allows an administrator to perform basic maintenance, including backups.



IDOL Admin allows administrators to quickly troubleshoot performance-related issues with an intuitive, visual dashboard

Languages (cont.)

- Mongolian
- Nahuatl
- Navajo
- Nepali
- Newari
- Norwegian
- Oriya
- Ossetian
- Panjabi
- Papiamentu
- Pashto
- Polish
- Portuguese
- Quechua
- Rhaeto-Romance (Romansh, Ladin and Friulian)
- Romanian
- Russian
- Sakha
- Sami
- Sanskrit
- Serbian
- Sesotho
- Sesotho sa Leboa/ Northern Sotho
- Sindhi
- Singhalese
- Slovak/Slovakian
- Slovenian
- Somali
- Sorbian
- Southern Ndebele
- Spanish
- Sranan
- Sundanese
- Swahili
- Swazi
- Swedish
- Syriac
- Tagalog
- Tahitian
- Tajik
- Tamil
- Tatar
- Telugu
- Thai
- Tibetan
- Tok Pisin
- Tongan
- Tsonga
- Tswana
- Turkish
- Turkmen
- Ukrainian
- Urdu
- Uyghur
- Uzbek
- Venda
- Vietnamese
- Waray-waray
- Welsh
- Wolof
- Xhosa
- Yiddish
- Yoruba
- Zulu

IDOL 10 offers over 500 functions

Feature	Capabilities provided
Conceptual search	IDOL 10 allows you to input a sentence or even an entire document as your query, since it can extract the main concepts and return the most relevant results. Searching by concept leads to the most comprehensive result set since it will also retrieve relevant documents that do not contain the query term(s) used.
Other search methods	IDOL 10 offers over 100 different search operators and modifiers to enhance the search experience. It supports keyword and Boolean search, parametric search, phonetic search, fuzzy search, wildcard search, federated search, geo-spatial search, social search, and many more.
Directed navigation	IDOL10 integrates information found in structured fields such as product name, size or manufacturer, with unstructured content, such as product description, allowing you to quickly narrow the search results to the subset of documents that contain the most relevant information.
Automatic categorization	IDOL 10 can organize the enterprise corpus of data according to pre-defined or dynamically generated categories for easy navigation. Existing legacy taxonomies can be either maintained or enriched with contextual understanding.
Automatic Query Guidance	IDOL 10 automatically groups search results into dynamically generated categories so that you can quickly narrow down your search set (e.g., a search for 'Madonna' would create clusters of the singer, the religious icon, and other relevant categories).
Federation	Queries may be federated to subordinate or exterior search engines.
Automatic Clustering	IDOL 10 can take a large set of data and automatically partition it so that similar information, even of varying formats, is clustered together. Each cluster represents a concept area, making it easier for you to identify inherent themes and emerging trends.
Advanced sentiment analysis	You can determine the degree to which a sentiment is positive, negative, or neutral for the entire content or a segment of the content. IDOL 10 uses both linguistic analysis and a statistical, pattern-based approach to derive sentiment. Currently offered for: Chinese, Dutch, English, French, German, Italian, Spanish, and Turkish.
Automatic Hyperlinking	IDOL 10 displays files (of varying formats, even audio and video) that are conceptually related to the file you are consuming. Links are generated in real time at the moment a document is viewed so you can delve deeper into the topic at hand.
Eduction (entity extraction)	IDOL 10 automatically identifies and extracts terms in documents that lend themselves to key fields such as the names of companies or people, locations, addresses, and telephone numbers. IDOL 10 offers hundreds of entities out-of-the-box across numerous languages, including multibyte languages such as Chinese, Japanese, Korean, and Thai. These languages pose particular challenges in entity extraction because words are not necessarily separated by spaces and they change meaning based on context. Conceptual entities can also be post-processed for validation purposes (e.g., is the credit card number valid?) or to customize information (e.g., replace SSNs with x's).
Personalization	IDOL 10 uses data that is based on your content consumption— browsing histories, content contributions, and interactions—with colleagues to construct an understanding of your interests and deliver more accurate, targeted results. You can also explicitly define your interests and train the search engine.
Expertise location	You can search for experts with IDOL 10. Instead of relying on seldom-updated profile information, you can use IDOL 10's conceptual picture of users to connect with experts.
Audio analytics	IDOL 10 combines phonetic and conceptual approaches to offer speech analysis functions with a high degree of accuracy. It accounts for speaker, dialect, and accent differences, background noise, codec variations, and speaking volume, and identifies topics being discussed, genders and emotional tones of speech, music, and linguistic origins of speakers.

Server Platforms

- Microsoft Windows XP/Vista/7/8, Microsoft Windows Server 2003, 2008, 2012 (x86 and x64)
- Redhat Enterprise Linux 5,6
- SUSE Linux 10,11
- Debian Linux 5,6
- CentOS 5,6
- Other POSIX compliant UNIX versions are available on request

Client Platforms

All client platforms are supported since IDOL 10 is based on standard HTTP and XML/JSON.

Feature	Capabilities provided
Video analytics	You can perform video search by concept, since IDOL 10 understands the content of the video without relying on metadata. It performs intelligent scene analysis, key framing, facial identification, OCR, logo recognition, barcodes detection, facial characteristics, skin types, clothing, security events (e.g., a gunshot), and many more.
Image analytics	IDOL 10 provides numerous image functions, including OCR, the ability to detect subtle patterns in images, identifying the same images from different angles, recognizing and analysing objects such as faces, bodies, gender, age range, expressions, and clothing, and can extract key data fields from scanned documents.
Native JSON output	IDOL 10 allows JSON output in addition to XML and XSL, lowering overhead for web developers.

IDOL 10 Platform Details

Scalable, distributed architecture	IDOL 10's distributed architecture easily supports environments ranging from a single server to thousands of servers. IDOL can dynamically scale for optimal performance while taking data volume, data sources, high availability, concurrent user count, query composition, frequency and response requirements into consideration. Both horizontal (scale-out) and vertical (scale-up) are possible with full support for virtualization.
Mapped security	IDOL 10 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 speeds query response times and ensures security entitlements are respected.
Language support	Fundamentally using a language-independent approach, IDOL 10 currently supports over 150 languages. IDOL 10 also provides localization tools to optimize performance such as stop lists and stemming.
High performance	IDOL 10 can support the most demanding enterprise requirements. It can process hundreds of millions of records every day, indexed at sub-second speed, and supports millions of users. IDOL currently analyzes over 50 petabytes of data in the HP Autonomy archives.
Connectivity	IDOL 10 can connect to over 400 system repositories, including the cloud.
Hadoop integration	IDOL 10 provides bi-directional connectors to popular big data systems such as Hadoop so that you can augment your investment without disruption. You can ingest Hadoop data into IDOL for advanced retrieval or push enterprise documents into Hadoop for MapReduce analysis.
Cloud applications data	IDOL 10 can search public content in the cloud and view the results along with those from the IDOL 10 index.
IDOL Admin	IDOL 10's administrative tool allows you to quickly answer performance-related questions with a visual dashboard. It displays resource allocation, query speed analysis, status, and progress of the index, errors, warnings in logs, and more.
Open architecture	IDOL 10 Server and its various modules and functions are easily integrated into other solutions via simple HTTP APIs and direct XML input and output methods. You can build custom solutions on top of the IDOL 10 platform.
File access	IDOL can analyze over 1000 file formats, including audio, video and image files.

About HP Autonomy

HP Autonomy is a global leader in software that processes human information, or unstructured data, including social media, email, video, audio, text, web pages, and more. Autonomy's powerful management and analytic tools for structured information together with its ability to extract meaning in real time from all forms of information, regardless of format, is a powerful tool for companies seeking to get the most out of their data. Autonomy's product portfolio helps power companies through enterprise search analytics, business process management and OEM operations. Autonomy also offers information governance solutions in areas such as eDiscovery, content management and compliance, as well as marketing solutions that help companies grow revenue, such as web content management, online marketing optimization and rich media management.

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

Learn more at
autonomy.com/idol10

Sign up for updates
hp.com/go/getupdated



Share with colleagues

© Copyright 2013 Hewlett-Packard Development Company, L.P. The information contained herein is subject to change without notice. The only warranties for HP products and services are set forth in the express warranty statements accompanying such products and services. Nothing herein should be construed as constituting an additional warranty. HP shall not be liable for technical or editorial errors or omissions contained herein.

Trademark acknowledgements if needed.

20130902_PI_B_HP_AUTN_IDOL10

