



Extending leadership in human information analytics

Autonomy IDOL 10



Benefits

- High-performance analytics – with enhancements in text, audio, video, and image analytics
- Social media analysis – with innovations in sentiment analysis
- An enriched Hadoop ecosystem – with various ways to combine IDOL and Hadoop capabilities in order to better analyze unstructured data
- Easy to use and administer – with a visual dashboard and simplified processes
- Reliable and robust – to increase performance while reducing down time

Next-generation information access platform purpose built for the era of information

We are in the midst of 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 be in one of these unstructured data types by 2015. Data growth is also accelerated by the ubiquity of mobile devices, which gives us "always on" applications and users, as well as new "information sensors" such as video recording cameras that are being widely adopted for surveillance, analysis, or entertainment purposes across an array of industries.

HP Autonomy's information access platform, Autonomy IDOL 10, has been built to address these modern challenges. This release reflects our renewed focus on delivering the industry's leading human information analytics capabilities while continuing to improve IDOL's ease of use and robustness. We have greatly simplified many administrative processes and made the platform even stronger for greater stability and performance, to meet the needs of the most demanding enterprises.

High-performance analytics

Topical sentiment analysis

For many organizations, the Internet can serve as a large focus group that operates free of charge. Insight and honest opinions can be found in people's unfiltered voices. Sentiment analysis technologies can help you quickly analyze text and determine positive or negative sentiment. But the problem is that people's reviews are often complex and multifaceted. They have mixed feelings, for instance, a harsh criticism can contain appreciation for one aspect of the product. Or unabashed admiration can include nit-picking of a flaw. Unfortunately, these subtleties are missed with traditional technologies.

To help you better identify people's endorsements and complaints, IDOL's sentiment analysis engine has been expanded to allow you to get a much more granular look at people's opinions within the context of a sentence. For example, consider a sentence that may be found on Twitter or a blog: "I stayed at Hotel XYZ last week, and though the mattresses were very nice, the service was awful." Ultimately, this shows mixed opinions about the XYZ hotel. In this scenario, traditional sentiment analysis technology would likely categorize this sentence as "neutral" or stack up points for both positive and negative sentiment, but it would do little to help XYZ isolate the real issue.

The new topic-level sentiment analysis allows decomposition and classification within a sentence to pull out specific topics (e.g., mattresses, service) and assign degrees of sentiment.

As IDOL 10 performs sentiment classification dynamically and in real time, you can realize immediate benefits. For instance, you can respond quickly to negative sentiment as opinions arise in the blogosphere, or adjust your marketing campaign to effect positive reception.

Personalized search experience

One major gap in today's search engines is the failure to acknowledge that each user is unique, and expects an answer catered to their specific interests. In fact, two people entering a query using the same keywords in an enterprise context are often seeking different documents based on their location, interests, job role, or project focus. A one-search-fits-all approach can leave users frustrated, with their goals unmet.

Autonomy IDOL 10 provides a highly personalized search experience that enables you to quickly find the information you need, when you need it. Using data based on the content you consume, content browsing histories, and content contributions and interactions with colleagues, IDOL constructs a complete profile of your interests to better analyze your intent and deliver more accurate, targeted results. You can also explicitly define your interests and train the search engine to understand and then deliver topics that matter to you.

IDOL 10 does not rank results simply on popularity, keyword matches, or other legacy algorithms, but automatically tailors search results to your purpose and specific query, making search a far more effective and pleasant experience.

Expanded conceptual entities

Imagine a series of documents on a certain topic that include names of experts scattered across the body of the content. Is there a quick way to isolate these names so they are easily shared? Or what about doing a search on "global warming" and being able to narrow the results by the specific places mentioned in the documents?

IDOL Education allows you to identify and extract entities (a word, phrase, or block of information) from text, based on a pattern you define. Sourcing from a vast library of entity grammars (patterns), IDOL Education 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 expands on this thousands-long list of out-of-the-box entities with new entity grammars that include medical drugs, airports, cars, and public figures. Each of these can then be flagged automatically to aid the search navigation process or populate database fields. Conceptual entities can also be post-processed so that they are validated (e.g. is the credit card number provided valid?) or customized (e.g. office location is formatted to output as "Headquarters" when it is in a certain zipcode).

Integration with Hadoop

HP Autonomy is embracing Hadoop as an important data store for today's organizations, and we are tightly integrating with Hadoop to enrich its unstructured content capabilities with IDOL's most advanced analytics capabilities.

Hadoop/HDFS connector

Our new connector to Hadoop/HDFS supports and augments the Hadoop ecosystem so that you can maximize your investment in this popular Big Data technology. Our connector offers an efficient, out-of-the-box solution for ingesting data stored in the Hadoop File Systems and provides a method of normalizing existing Hadoop data ready for the Map phase.

Data ingest into Hadoop

HP Autonomy can use its standing connector infrastructure along with the Hadoop connector to push enterprise documents into Hadoop for MapReduce analysis. For instance, we can take a collection of chat data, database extract, or log data from your enterprise file system to ingest data into Hadoop. Or we can help clients archive content from data stores being retired (e.g. old document management systems) and move it to Hadoop.

Analysis of data already in Hadoop

IDOL technology can be embedded in a standard Hadoop instance to perform a variety of

advanced analytics tasks such as extracting file content and metadata, performing sentiment analysis, and classifying or clustering documents. By using IDOL components in an existing environment and a programming framework (MapReduce), you can reduce coding and implementation costs while performing complex analysis that improves business outcomes.

Index Hadoop content with IDOL

You can create an index of Hadoop documents in IDOL using our Hadoop connector. With this approach, you can perform much faster and more sophisticated analysis than MapReduce for content stored in Hadoop.

Easier than ever to consume and administer

With IDOL 10, administrators can quickly troubleshoot performance-related issues with an intuitive, visual dashboard so you can remain productive. These questions include:

- Why is my query slower than expected?
- Is my license going to expire soon?
- Have I correctly configured IDOL?
- What is the status of indexing content?
- What is causing the usage of system resources (RAM) and what can I do to reduce that?

Answers to these questions can be easily viewed in a graphical presentation. 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

Enjoy fast performance and avoid downtime

Dynamic capacity expansion

The amount of information in your enterprise is expanding at an increasingly faster rate. To accommodate the growth of the IDOL index, an administrator may need to consider adding more IDOL engines, which may require re-indexing existing documents. We have developed a new distribution method that allows you to dynamically expand your capacity and add new engines without re-indexing. The result is improved performance and no downtime—a must for all mission-critical applications.

Self-diagnosis of indices

IDOL 10 helps you recover intelligently from crashes and failures with automatic testing of the integrity of indices upon restart. We have added a wide range of validation options to check against corruption in the event of a hardware failure or system crash. Sub-indices can be validated individually and the engine can be configured to automatically check the status of the indices at start up.

Additional advantages

In addition to the above features, IDOL 10 also provides additional platform enhancements, including audio and image analytics, security, performance, robustness, and customer-reported issues.

Server Platforms

IDOL 10 is supported on the following 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

Key features of IDOL 10:

| Core functionality | |
|---|---|
| Feature | Capabilities provided |
| Topical sentiment analysis | Analyze sentiment at a granular level, even detecting different sentiments within a sentence. Additional language features: Chinese, French, German, and Italian |
| Entity extraction | Improve search by automatically extracting 1000s of conceptual entities; added new education libraries and included education server post-processing |
| Hadoop connector | Ingest Hadoop data into IDOL for advanced analytics, extract metadata and conduct analytics for files stored in Hadoop, push enterprise documents into Hadoop (chat data, ODBC, documents) for MapReduce analysis |
| Personalized ranking | Deliver search results personalized by the user's profile (implicit/explicit) |
| Improved audio and image analysis | Detect barcodes, facial characteristics, skin types, clothing, security events (e.g., a gunshot) |
| Phonetic (sounds-like) search | Improve accuracy by up to 30% more than next-best algorithm |
| Native JSON output | Allow JSON output in addition to XML and XSL, and lower overhead for web developers. |
| Enhanced sorting | Sort by frequently-used business rules |
| Platform enhancements | |
| IDOL Admin | Quickly answer performance-related questions with new visual dashboard. Display resource allocation, query speed analysis, status and progress of the index, errors, and warnings in logs, etc. |
| Dynamic capacity expansion | Expand capacity and add new engines without the need to re-index, resulting in improved performance and no downtime (useful when running out of space or adding new sources) |
| Improved data transmission | Increase your indexing speed by as much as 47x with improved data transmission. Administrator can fine-tune TCP settings to optimal values, resulting in increased performance, especially for geographically-distributed data centers. |
| Direct audio and image processing | Process audio and image data like you would with any other file type and extract meaning from these files with minimal setup |
| Expanded SharePoint and Exchange connectors | Offer different ways to connect to SharePoint (use native SharePoint web services) and Exchange (does not rely on MAPI) |
| Self-diagnosis of indices | Recover intelligently from crashes and failures with automatic testing of the integrity of indices upon restart, and eliminate the need for re-indexing |
| Secure deletion of content | Delete confidential documents without possibility to un-delete |
| Document tracking | Prevent the loss of documents during the indexing process by pushing errored documents into special content engines for review |
| Index term statistics | Improve performance and search output by allowing system administrators to detect unusual documents, or documents that were programmatically created to contain millions of "garbage" terms |
| Configuration auto-validation | Check configuration files to detect errors in file parameters |

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 and web pages, etc. 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.

Sign up for updates
hp.com/go/getupdated



Share with colleagues

Copyright © 2013 HP Autonomy. All rights reserved. Other trademarks are registered trademarks and the properties of their respective owners. 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.

autonomy.com

20130426_PI_DS_HP_Autonomy_IDOL10

