IDOL

Unified text analytics, speech analytics and video analytics
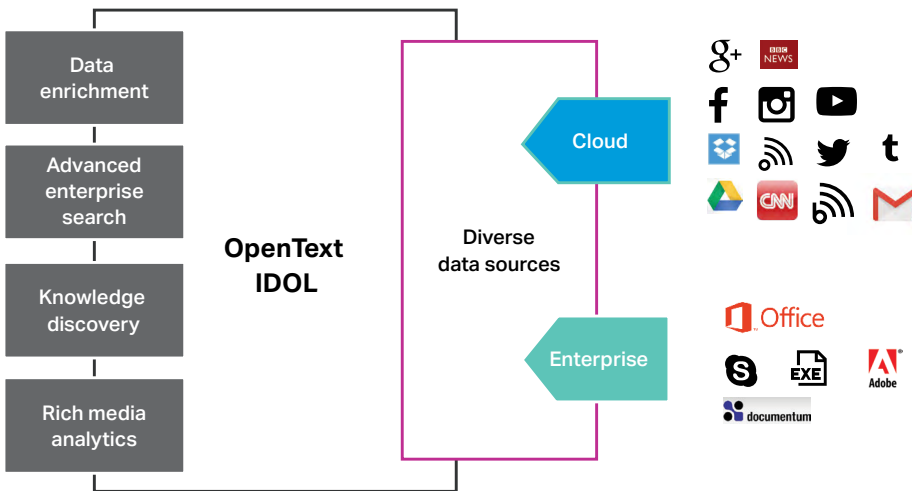

Figure 1. Typical use cases for IDOL

IDOL at a Glance:

- **Harness virtually any data:** With connectors for 150+ sources and support for over 1000 data types, you can access virtually any data inside and outside your firewall, and index data without relocation or disruption.

- **Unified text analytics, speech analytics and video analytics:** Get the actionable insights by unlocking patterns, trends, and relationships across siloed unstructured data repositories by using the latest innovations in machine learning and Deep Neural Networks.

- **Enterprise grade security:** Make sure the right data gets to the right people. Preserve and stay up-to-date on security entitlements without compromising performance by using strong data access protection.

Product Highlights

OpenText™ IDOL provides unified text analytics, speech analytics and video analytics with support for over 1000 data formats. It enables out-of-the-box access to 150 data repositories behind and beyond your firewall, (for example, Documentum, Dropbox, and such) as well as indexes data without relocation and disruption. Built on proven world-class technology and innovations such as machine learning and deep neural networks, IDOL unlocks hidden insights by revealing trends, patterns, and relationships. With IDOL, you gain an in-depth understanding of user profiles and actions to personalize knowledge delivery. In information search and knowledge discovery, IDOL natural language question answering transforms the end-user experience. It enables simple and contextually relevant dialogues between humans and machines.

With modular offerings of hundreds of advanced analytics functions, as well as an open and scalable architecture for easy embedding and third-party integration, IDOL lends itself to supporting diverse use cases spanning a broad spectrum of industries.

Protecting confidentiality within an organization is critical to effective operations. IDOL mapped security enables preservation of security entitlements. Synchronized updates of such entitlements enable users to access the right information while simultaneously safeguarding sensitive data.

Key Benefits

Understand and Act on Human Data
Data enrichment is about augmentation with other relevant data, such as metadata. For example, IDOL can extract company names from tweets, associate the tweets with the extracted names, and make the tweets filterable by company name.

Advanced enterprise search is about providing enterprise search results based upon relevant concepts associated with the search terms. This goes beyond simple key-word search. A conceptual enterprise search example—if you search for Apple, you may see results associated with Apple, Samsung, and Microsoft because IDOL understands that these companies are related, in that they're in the same industry and address similar markets.

Knowledge discovery liberates users from having to know what questions to ask beforehand. IDOL builds on machine learning and deep neural network algorithms to recognize patterns, trends, and relationships hidden within the data and let the data tell the story. For example, IDOL can analyze customers' tweets and call center logs to reveal root causes as to why a product may be underperforming.

Rich media analytics allows users to incorporate video, image, and audio content to gain complete data insights. For example, in addition to text analytics of social media, a marketer can also monitor and analyze broadcast media for logos, on-screen text, and speech.
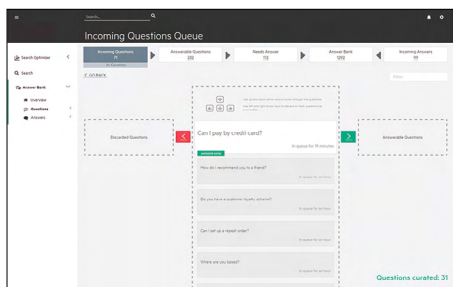

Figure 2. Natural Language Question Answering Admin UI simplifies question/answer pair curation

Key Features
Dynamic Corpus enables unprecedented information access with dynamic data-led discovery of relevant sources including the Dark Web, which might otherwise be inaccessible.

IDOL Ingest for NiFi enables IDOL's robust ingest capabilities to be used as components/ processes in the NiFi framework. It significantly simplifies configuration and monitoring of the ingest chain for easier scalability, fault tolerant and document tracking.

Encryption uses existing AES/FIPS cryptography to encrypt indices and document, safeguarding against unauthorized access and replication.

PII (inclusive of GDPR) grammars are now available for 36 countries (28 EU countries together with Australia, Canada, Iceland, Lichtenstein, New Zealand, Norway, Turkey, United States), which enable explicit targeting of Personable Identifiable Information (PII) to simplify compliance.

Natural language question answering accepts queries in natural language. A variety of question examples may include "how do I turn off roaming on iPhone 8?", "what was the earnings per share (EPS) of APPL (Apple Inc.) in Q3 2019?", or "what did Apple's CEO say about virtual reality?"

IDOL can be configured to create a more natural system of customer services. Rather than a single search box, an IM-style interface allows a user to enter an initial question or problem description in natural language. In many cases, the problem can be answered directly and a single answer or solution provided, but in many other cases a conversation between system and the user is activated to gather more information on how to diagnose and solve the problem.

For ongoing refinement of an answers' availability, accuracy, and relevancy, administrators can use the intuitive GUI based tool for improving the efficiency of manual curation of question/answer pairs.

Automatic query guidance groups search results into dynamically generated categories so you can quickly narrow down a search set (for example, a search for "Madonna" would create clusters of the singer, the religious icon, and other relevant categories).

Hyperlinking displays links to files of varying formats (including audio and video) that are conceptually related to the document you are currently consuming. If you are reading an article about a competitor, IDOL may display links to the company's latest ad campaign, an audio interview with a CEO, internal competitive battle cards, and such. Links are generated in real time when a document is viewed so you can delve deeper into the topic at hand.

Retrieval concept allows you to input a sentence or even an entire document as your query because it can extract the main concepts without relying on key words. Searching by concept leads to the most comprehensive result set because it will also retrieve relevant documents that do not contain the actual query term(s) used.

Personalization constructs an understanding of your interests and skillset to deliver more accurate, targeted enterprise search, and knowledge discovery results based on content consumption including browsing histories, content contributions, and interactions. You can also explicitly define your interests and train the search engine.

Expertise locator identifies experts based on their content activities and social cues instead of relying on outdated profile information entered years ago.

Categorization and channels organize the enterprise corpus of data according to predefined or dynamically generated categories for easy navigation. Existing legacy taxonomies can be either maintained or enriched with a contextual understanding.

Eduction 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 offers hundreds of entity grammars out of the box across numerous languages.

Clustering takes a large set of data and automatically partitions it so that similar information, even in varying data formats, is clustered together. Each cluster represents a concept area, making it easier for you to identify inherent themes and emerging trends.

Sentiment analysis determines the degree to which a given text's sentiment is positive, negative, or neutral for the entire content or a segment of the content. IDOL uses both linguistic analysis and a statistical, pattern-based approach to derive sentiment. Currently, it is offered in Arabic, Chinese, Czech, English, French, German, Italian, Polish, Portuguese, Russian, Spanish, and Turkish.

Visualization is provided by Advanced Find, an end-user graphical user interface with functionality that includes a topic map to highlight key concepts. It includes a sunburst diagram to enable easy filtering based on extracted entities (for exam- people, place, company, and more). The result is set as a comparison to examine how a change of search parameter may impact the outcome. Advanced Find offers a rich contextual view where the query result includes not only the document itself but also shows metadata and other relevant information such as documents by the same author or documents from around the same period.

Customizable search interface, Find, is an open source tool from OpenText™. It is an extensible enterprise search user interface, which now ships with IDOL for our customer's convenience. It is designed to be a highly flexible foundation upon which custom-made applications can be based. The quick-to-start and easy-to-use enterprise search interface enables almost instantaneous search productivity while accelerating the development of applications requiring intelligent search capabilities.

Knowledge graph uses advanced graph analytics technology to discover relationships

between entities that lead to richer and more impactful knowledge discovery. Find out who knows whom, who knows what, what shared common traits exist among your important customers, and more.

Connectors provide access to enterprise content management system (CMS) and mail applications in addition to supporting open protocols and cloud-based systems to provide a wide data collection for index generation.

Security capabilities include synchronized security entitlement preservation, authentication, access control, and user homogenization.

Open and robust architecture components allow for easy integration, high availability, and scalability options.

System Administration
- User/role management
- Memory and thread usage monitoring
- Service monitoring and control
- IDOL database creation monitoring
- Module configuration
- Host/service discovery
- Configuration validation
- Document tracking
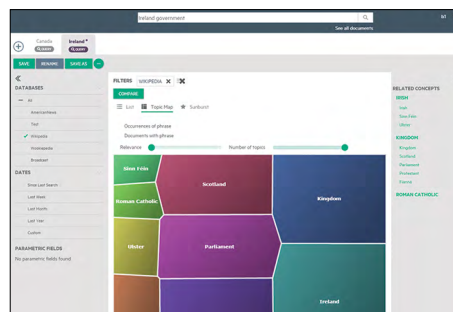- Search covers 150 languages

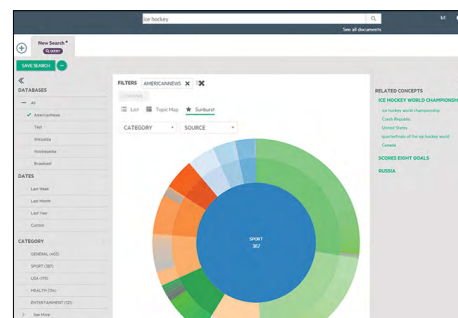### Advanced Find

Figure 3. Key concepts with easy drill-down

Figure 4. Filter based on extracted entities (people, place, company) from data
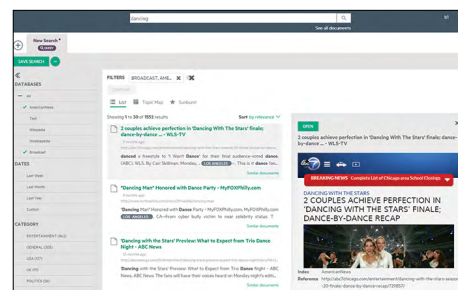

Figure 5. Direct preview of original information from search results
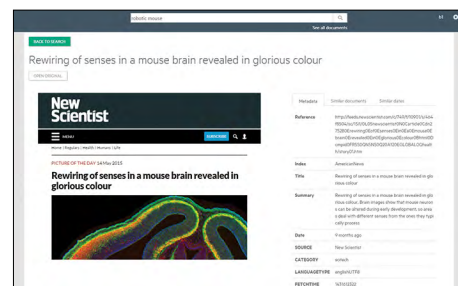

Figure 6. Detailed summary of document's metadata, list of similar documents, and documents with similar dates

The IDOL Admin dashboard allows administrators to troubleshoot performance-related issues quickly.
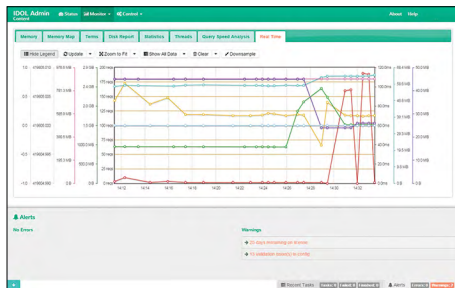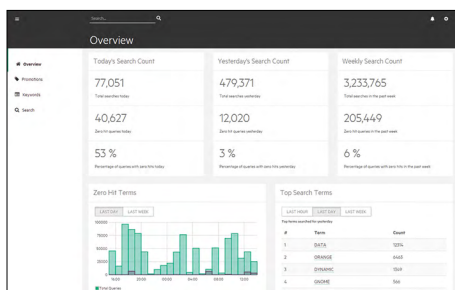
Figure 7. IDOL Admin dashboard

Figure 8. IDOL Data Admin

The IDOL Data Admin provides business administrators with a dashboard view of various statistics so they can refine to improve the search experience.

Platforms Supported

The platforms supported vary upon the individual components. This is an inevitable result of the complexity of integrating with software supplied by multiple vendors. Provide your sales team with your requirements for definitive guidance on supported platforms.

Product Tiers

IDOL Express Search

This is licensed by the number of documents on an unlimited number of CPUs and for an unlimited number of users, operating across a single cluster. In IDOL Express Search, if a document is stored more than once, each instance is included in the count separately.

The features and functions included in IDOL Express Search are as follows:

- Functions
  - Retrieval—Lite
  - Retrieval—Concept
  - Retrieval—Advanced
  - Retrieval—Parametric
  - Hyperlinking
  - Summarization
  - Dynamic Thesaurus
  - Spelling Correction
  - Automatic Language Detection
  - Highlighting
- Infrastructure tools
  - Distributed Index Handler
  - Distributed Action Handler
  - Application Builder (ACI API)
- System administration tools
  - IDOL Admin
  - IDOL Site Admin
- Data administration tools
  - IDOL Data Admin
- End-user tools
  - Find
  - BIFHI
- Connectors
  - IDOL Express Search includes all IDOL standard connectors

IDOL Premium Search

IDOL Premium Search is licensed by the number of documents on an unlimited number of CPUs and for an unlimited number of users, operating across a single cluster. If a document is stored more than once, each instance is included in the count separately. The features and functionality in IDOL Premium Search include those in IDOL Express Search and the following:

- Functions
  - Generic mapped security (IAS)
  - Personalization package (alerting, mailing, agents, and more)
  - Expertise package (collaboration, expertise, profiling, and more)
  - Automatic Query Guidance (AQG)
  - Agent Boolean

Other functionality can be added to IDOL Premium Search (see the following IDOL add-on modules.)

- Connectors
  - IDOL Premium Search includes all IDOL standard connectors

IDOL Ultimate

IDOL Ultimate is licensed by the number of documents on an unlimited number of CPUs and for an unlimited number of users operating across a single cluster. If a document is stored more than once, each instance is included in the count separately. The features and functionality in IDOL Ultimate include those in IDOL Premium Search listed in the preceding section, along with the following:

- Functions
  - Sentiment analysis
  - Eduction
  - Natural Language Question Answering
  - Clustering
  - Knowledge graph analytics
  - Taxonomy generation
  - Categorization and channels
  - Geospatial
- Business administration tools
  - IDOL Collaborative Classifier
- Connectors
  - IDOL Ultimate includes all IDOL standard connectors

IDOL Premium Search Add-Ons

IDOL Premium add-on modules provide additional functionality and can be applied to IDOL Premium Search. They are licensed by the number of documents and tie in with a purchase of IDOL Premium Search software. If a document is stored more than once, each instance is included in the count separately.

- **Available add-on modules**
  - Sentiment analysis
  - Conversational Capabilities
  - Eduction
  - Clustering

IDOL Dynamic Corpus

IDOL Dynamic Corpus is a separately licensed module that enables a connector with dynamic data-led discovery to deliver information of high relevancy. Dynamic Corpus applies to Web connector.

IDOL Connectors

IDOL Standard Connectors

IDOL standard connectors are included with IDOL Express Search, IDOL Premium Search, and IDOL Ultimate. They are not separately licensable. IDOL standard connectors include the ability to connect to and access data from repositories. The following connectors are IDOL standard connectors:

- File system (all platforms)
- ODBC
- Web

IDOL Social Media Connectors

These connectors are add-on modules that can be applied to IDOL Premium Search and IDOL Ultimate. They are licensed by a number of IDOL Social Media Connectors and a license key is for a specified number of social media connectors. To modify which specific connectors are licensed, you must obtain a new license key. Modifications to connector license keys can be done once per month.

IDOL Social Media Connectors include the ability to connect to and access data from repositories, on a polling basis. The following connectors are IDOL Social Media Connectors:

- Facebook
- Amazon
- Twitter
- LinkedIn
- YouTube
- Atom
- Quickr
- Weibo
- Yahoo

IDOL Advanced Connectors

IDOL Advanced Connectors are add-on modules that can be applied to IDOL Premium Search and IDOL Ultimate. They are licensed by a number of Advanced Connectors and a license key is for a specified number of Advanced Connectors. To modify which specific connectors are licensed, you must obtain a new license key. Modifications to connector license keys can be done once per month. Advanced Connectors include the ability to connect to and access data from repositories.

The following connectors are IDOL Advanced Connectors:

- Box
- Chatter
- Confluence
- DataSift
- Documentum
- Dropbox
- Evernote
- Exchange
- Exchange Web Services
- IMAP
- Lotus Notes
- Lync Archiving
- Lync Group Chat
- OCS
- OneDrive
- Oracle
- Oracle UCM
- POP3
- RSS
- Salesforce
- SAP NetWeaver
- SharePoint
- SharePoint Remote
- Skype for Business
- SourceOne
- OpenText StoreAll
- Twitter streaming

IDOL Plus Connectors

IDOL Plus Connectors are add-on modules that can be applied to IDOL Premium Search and IDOL Ultimate. They are licensed by a number of IDOL Plus Connectors and a license key is for a specified number of IDOL Plus Connectors. To modify which specific connectors are licensed, you must obtain a new license key. Modifications to connector license keys can be done no more often than once per month.

IDOL Plus Connectors include the ability to connect to and access data from repositories. The following connectors are IDOL Plus Connectors:

- Alfresco
- Drupal
- Documentum PLUS(+)
- Ektron
- ENOVIA
- Enterprise Vault
- eRoom
- Exchange PLUS(+)
- FileNet P8 PLUS(+)

- Hadoop, IBM Content Manager
- IBM Seedlist
- JIRA
- Jive
- Lotus Notes PLUS(+)
- OpenText/Hummingbird
- SAP NetWeaver
- SharePoint PLUS(+)

IDOL Plus Connectors marked (+) also include one or more advanced features such as:

- **View:** View a document in native or near-native view
- **Collect:** Retrieve documents and send them to a destination, such as a shared folder
- **Hold:** Make a set of documents unchangeable in the source repository
- **Release:** Feature provides the opposite capability of Hold
- **Delete:** Delete, unlink, and remove a set of documents from a repository
- **Update:** Change repository metadata about a set of documents (created time, author, and others)
- **Insert:** Add a document into a repository

MMAP (Media Management and Analytics Platform)

The media analytics and management platform provides all components necessary to facilitate analysis, management, and visualization of large volumes of rich media allowing:

- Video analysis and enrichment
- Real-time monitoring and alerting
- Search and retrieval
- Recording and review from a wide array of sources:
- Dedicated streaming sources:
  - CCTV cameras
  - Mobile devices

- File-based rich-media archives:
  - Voice call recordings
  - Enterprise data lakes
  - Digital asset management (DAM)
- Public domain sources
  - Broadcast TV and radio
  - Online video/audio archives
- Social media sites
  - Facebook
  - Instagram
  - Twitter
  - YouTube

A part of OpenText IDOL, MMAP supports:

- **Usability**—highly intuitive and simple to use
- **Availability**—cloud, on premises, and on mobile devices
- **Performance**—industry-leading accuracy and speed
- **Enterprise**—robust scalable architecture
- **Integration**—seamlessly integrated within IDOL for comprehensive analytics across unstructured, semi-structured, and structured data sources
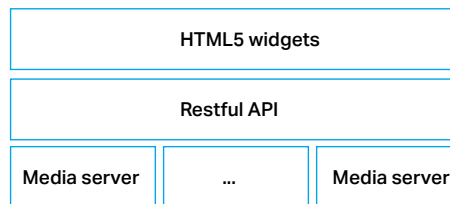
Media Servers

HTML5 widgets

Restful API

Media server	...	Media server

Figure 9. MMAP Components

Audio Analytics

Using Web APIs the server provides analysis of audio files and streams. Integrated with IDOL Connector Framework Server (CFS) provides indexing of analysis direct into IDOL. Each server is licensed per channel of ingest and for that channel any combination of the media analytics can be applied.

Visual Analytics

Using Web APIs the server provides analysis and encoding of the visual components of image files, video files, and video streams, outputting analysis results to data repositories. Integration with IDOL Connector Framework Server (CFS) provides indexing of analysis direct into IDOL. Each server is licensed per channel of ingest and for that channel any combination of the media analytics can be applied. When combined with speech (audio) server licenses, visual server provides visual and audio analysis in one seamless package.

Surveillance Analytics

Using Web APIs the server provides dedicated video analysis and encoding for IP CCTV surveillance camera streams, outputting analysis results to data repositories. Integration with IDOL Connector Framework Server (CFS) provides indexing of analysis direct into IDOL. Each server is licensed per camera IP stream and for that IP stream any combination of the analytics can be applied.

Glossary

Cluster: A set of instances created on one or more servers for the express purpose of improving the capacity, performance, or availability characteristics of a single instance

Concurrent media asset: Simultaneous use of more than one media asset at any point in time

Core: A part of a CPU that executes a single stream of compiled instruction code

CPU: A system in a single integrated circuit package with one or more discrete processing cores

Document: A text unit (and its associated metadata) that is independently returnable as a search result

Audio Analytics

Type	Function	Description
Speech	Speech to text	Converts spoken speech to a text transcript of the most likely words
Speech	Phonetic phrase search	Converts spoken speech to a phonetic index, which can be searched against
Speech	Speaker segmentation and identification	Identifies speakers in spoken speech
Speech	Spoken language identification	Identifies the language spoken
Speech	Transcript alignment	Aligns a given text transcript with an audio file producing time stamps for all words
Audio	Audio quality and classification	Classifies audio segments as music, noise, or speech, as well as giving details on the audio quality
Audio	Audio fingerprint identification	Allows creation of an audio database for identifying audio segments
Audio	Audio security	Identifies common security threats from audio captured
Model	Language customization	Allows customization of language models used in speech-to-text operations
Model	Acoustic adaptation	Allows adaptation of acoustic models used in speech-to-text, phonetic-phrase-search operations

Visual Analytics

Type	Function	Description
Image	Barcode	Detect and read QR codes
Video/Image	FaceDetect	Detect faces
Video/Image	Demographics	Obtain demographic information such as age, gender, and ethnicity for detected faces
Video/Image	FaceRecognize	Run face recognition on detected faces
Video/Image	FaceState	Obtain additional information, such as facial expression, about detected faces
Video	Keyframe	Identify keyframes
Video/Image	Object	Recognize known objects in video
Video/Image	ObjectClas	Recognize known object classes in video
Video/Image	OCR	Run intelligent scene analysis to identify important
Video	SceneAnalysis	Run intelligent scene analysis to identify important events
Video/Image	NumberPlate	Detect and read license plates on vehicles, including the ability to Identify color and make of vehicle

Surveillance Analytics

Source	Function	Description
Camera	SceneAnalysis	Run intelligent scene analysis to identify important events
Camera	NumberPlate	Detect and read license plates on vehicles, including the ability to Identify color and make of vehicle
Camera	FaceDetect	Detect faces
Camera	Demographics	Obtain demographic information such as age, gender, and ethnicity for detected faces
Camera	FaceRecognize	Run face recognition on detected faces

E-LTU and E-Media: Products, which are only delivered electronically; as such, any reference to FOB destination or delivery methods that are stated on your purchase order other than electronic shall be null and void with respect to these E-LTU or E-media products

GB: Gigabytes, which designates the amount of physical capacity that can be managed

IDOL raw data size: The uncompressed size of all documents, including text and metadata that are stored in a node table excluding any deleted documents that have not been purged

Ingest channel: A single input to the analytics server that can handle one external source of rich media at a time

Instance: Each implementation of the application installed on a server

LTU: License to use

Media asset: A single file containing either video, audio, or image content

Rich media: Video, audio, or image content either as a file or continuous stream via IP

Server or SVR: Means any designated computer system in which an instance or instances of the software are installed

TB: Terabytes, which designates the amount of physical capacity that can be managed

Term license to use or Term LTU: A software LTU, which indicates in its license description that the license is valid for a specific period of time such as one month (1M), one year (1Y), and the like; term LTUs are not perpetual licenses

Term support: A fixed period support offering that is only valid during the period of the associated Term LTU

opentext™

Unlimited or Unl: Without restrictions in terms of a number of systems, devices, or media, depending on the context

User: A user whose use is restricted to the type of software that is being licensed

Learn more at
www.microfocus.com/en-us/portfolio/unstructured-data-analytics-platform

www.opentext.com

opentext™