

## DATA SHEET

# OpenText Knowledge Discovery

Unified text analytics, speech analytics, and video analytics

### Knowledge Discovery at a glance

- **Harness virtually any data:** With connectors for more than 160 sources and support for more than 2,000 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 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.

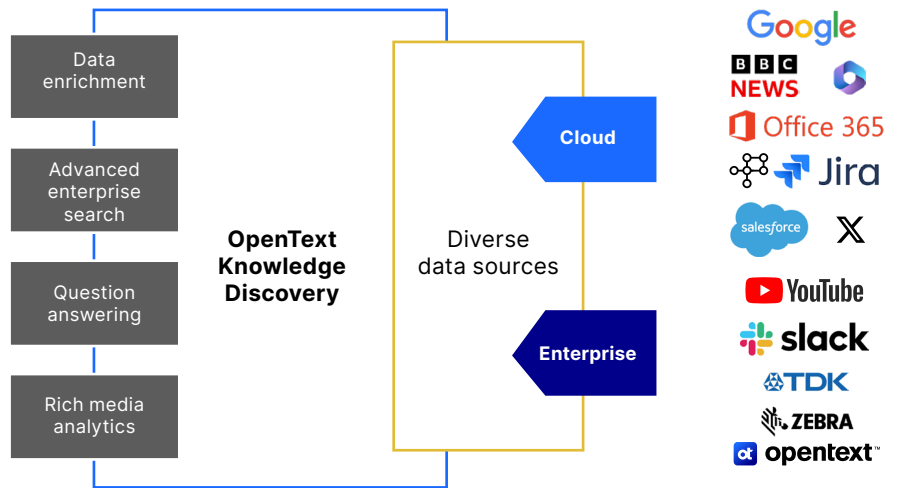


Figure 1. Typical use cases for Knowledge Discovery

### Product highlights

**OpenText™ Knowledge Discovery** (formerly IDOL) provides unified text analytics, speech analytics, and video analytics with support for more than 2,000 data formats. It enables out-of-the-box access to more than 160 data repositories behind and beyond your firewall, (for example, OpenText™ Extended ECM, OpenText™ Documentum™, Microsoft, etc.) indexing data without relocation and disruption.

Knowledge Discovery is built on proven world-class technology and innovations, such as machine learning and deep neural networks. The solution unlocks hidden insights by revealing trends, patterns, and relationships. You gain an in-depth understanding of user profiles and actions to personalize knowledge delivery. Knowledge Discovery's 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, Knowledge Discovery lends itself to supporting diverse use cases spanning a broad spectrum of industries.

Protecting confidentiality within an organization is critical to effective operations. Security mapped with the solution 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, you 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 results based upon relevant concepts associated with the search terms. This goes beyond simple keyword search. A conceptual enterprise search example—if you search for Apple, you may see results associated with Apple, Samsung, and Microsoft because Knowledge Discovery 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 before-hand. It 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, it can analyze customers' tweets and call center logs to reveal root causes as to why a product may be underperforming.

**Rich media analytics** allow 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.

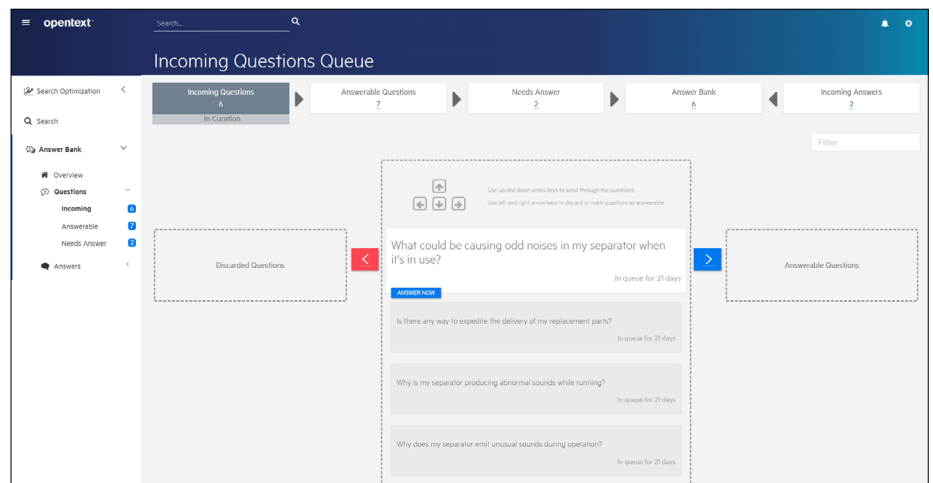


Figure 2. Natural language question answering Data Admin simplifies question/answer pair curation

## Key features

**Knowledge Discovery** enables unprecedented information access with dynamic data-led discovery of relevant sources, including the dark web, which might otherwise be inaccessible.

**Ingest workflow** enables robust ingest capabilities to be used as components/processes. It significantly simplifies configuration and monitoring of the ingest chain for easier scalability, fault tolerance, 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 personally 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?”

Knowledge Discovery 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 with 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, Knowledge Discovery 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. Results are provided 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. Knowledge Discovery 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.

## System administration

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

**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. Knowledge Discovery 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** functionality includes a topic map to highlight key concepts. It includes a sunburst diagram to enable easy filtering based on extracted entities (for example: 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.

The Knowledge Discovery UI 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.

## Advanced interfaces

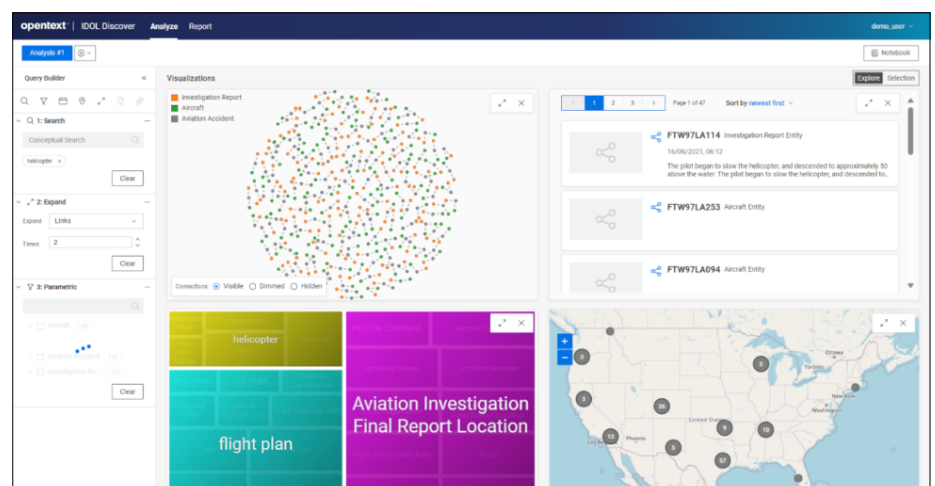


Figure 3. Key concepts with easy drill-down

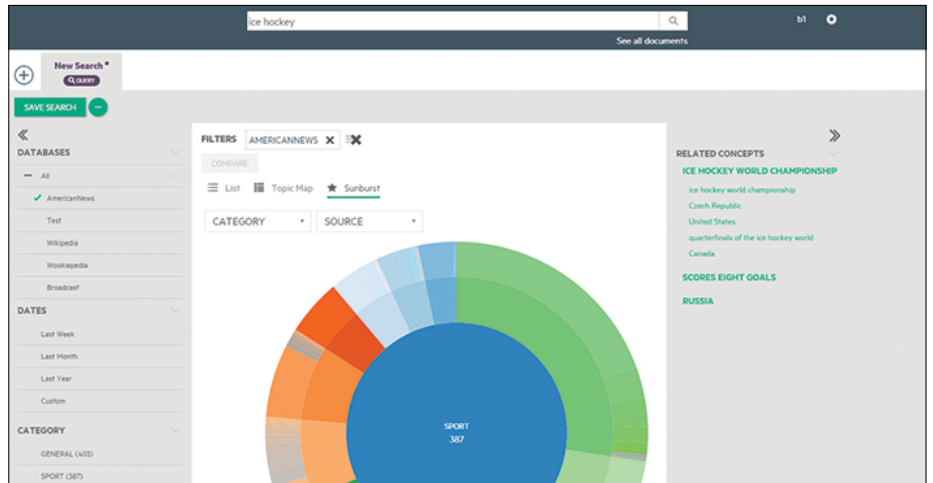


Figure 4. Filter based on extracted entities (people, place, company) from data

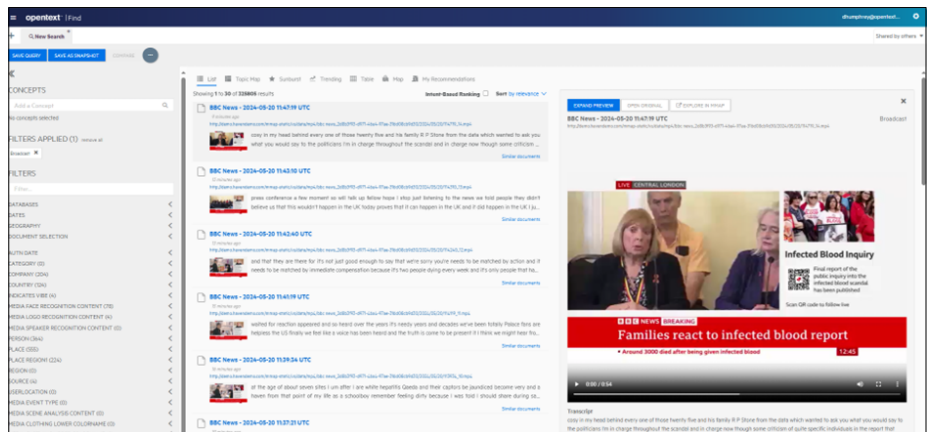


Figure 5. Direct preview of original information from search results

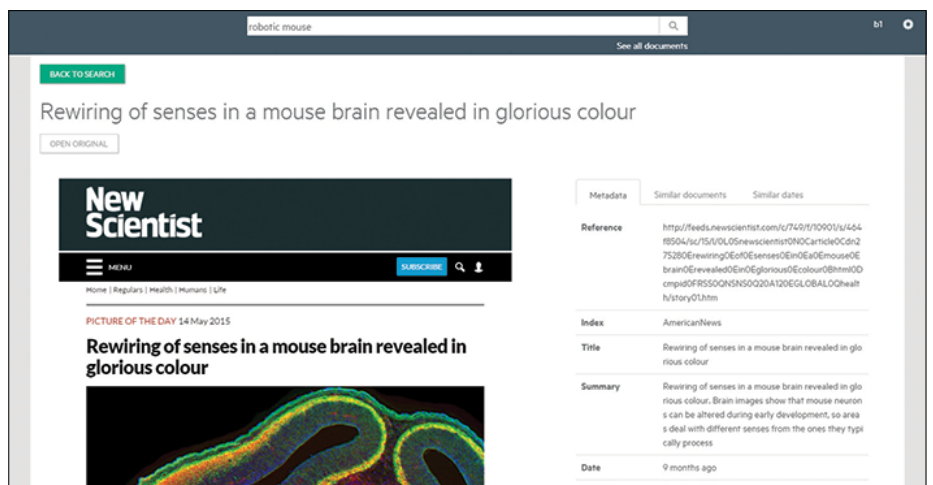


Figure 6. Detailed summary of document's metadata, list of similar documents, and documents with similar dates

The Knowledge Discovery Admin dashboard allows administrators to troubleshoot performance-related issues quickly.

The Knowledge Discovery Admin dashboard allows administrators to troubleshoot performance-related issues quickly.

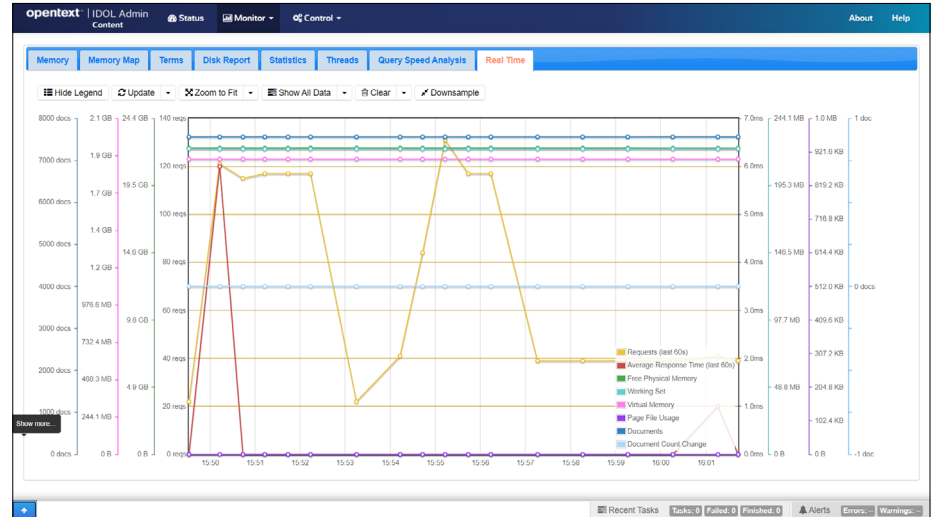


Figure 7. Admin dashboard

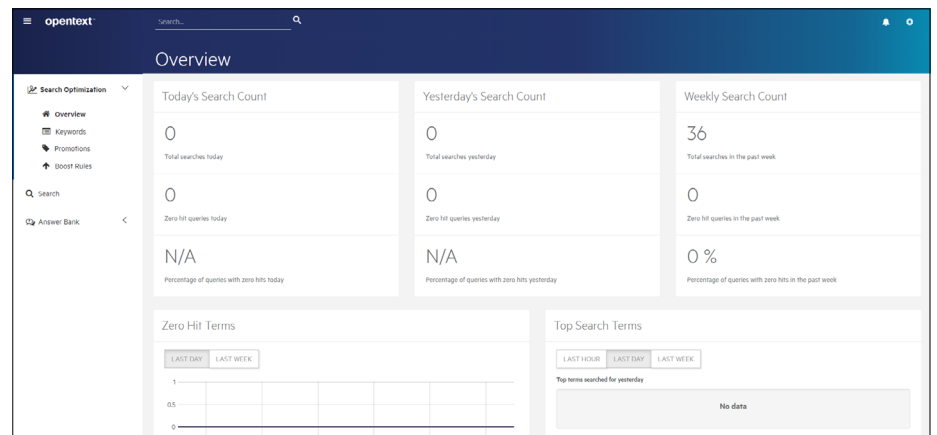


Figure 8. Data Admin interface

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

## Product capabilities

Knowledge Discovery features and functionality include:

### Functions

- Agent Boolean
- Automatic language detection
- Automatic query guidance
- Categorization and channels
- Clustering
- Conversational capabilities
- Dynamic corpus
- Dynamic thesaurus
- Education
- Expertise, such as collaboration, expertise, profiling, and more
- Geospatial
- Generic mapped security
- Highlighting
- Hyperlinking
- Knowledge graph analytics
- Natural language question answering
- Personalization package (alerting, mailing, agents, and more)
- Retrieval—concept
- Retrieval—parametric
- Retrieval—vector
- Sentiment analysis
- Spelling correction
- Summarization
- Taxonomy generation

### Advanced features

- **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

### Infrastructure tools

- Distributed index handler
- Distributed action handler
- Application builder (ACI API)

### System and data administration tools

- Component admin
- System site admin
- Data admin

### End-user tools

- Teams bot
- Exchange bot

## Knowledge Discovery - connectors

160+ connectors including:

### • OpenText Content Cloud

- OpenText™ Extended ECM
- OpenText™ Core Content
- OpenText™ Documentum™
- OpenText™ InfoArchive

### • Microsoft®

- Exchange®
- Teams®
- SharePoint®

### • Google

- Workspace™
- YouTube™
- Drive™

### • Salesforce®

### • Amazon S3

### • Microsoft® Azure Blob

### • Box

### • Jira

### • File System

### • Social Media

- Facebook®
- Weibo
- LinkedIn
- TikTok
- X (formerly Twitter™)
- Telegram



## Rich media analytics

Knowledge Discovery media analytics 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
  - X (formerly Twitter)
  - YouTube

Knowledge Discovery media analytics 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 for comprehensive analytics across unstructured, semi-structured, and structured data sources

## Media servers

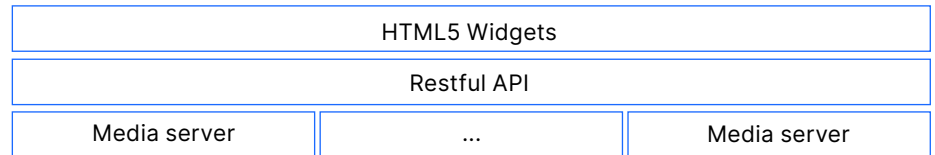


Figure 9. Media analytics components

### Audio Analytics

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

## Knowledge Discovery 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	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

## Knowledge Discovery 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
NumberPlate	NumberPlate	Detect and read license plates on vehicles, including the ability to identify color and make of vehicle

[Learn more](#)

### About OpenText

OpenText, The Information Company, enables organizations to gain insight through market leading information management solutions, on premises or in the cloud. For more information about OpenText (NASDAQ: OTEX, TSX: OTEX) visit: [opentext.com](https://www.opentext.com).

### Connect with us:

- [OpenText CEO Mark Barrenechea's blog](#)
- [X \(formerly Twitter\)](#) | [LinkedIn](#)