opentext™

Powered By IDOL

OEM SDKs for unstructured data analytics

OpenText™ IDOL SDKs provide unstructured data analytics to leading edge software developers and service providers, reducing engineering risk, time to market, development costs, and ongoing maintenance by leveraging IDOL's unparalleled history in artificial intelligence and machine learning.

Data Collection

Maximize the amount of data your application can access. Retrieve and aggregate data, including repository metadata, from local, remote, and cloud repositories, making it accessible to downstream processing.
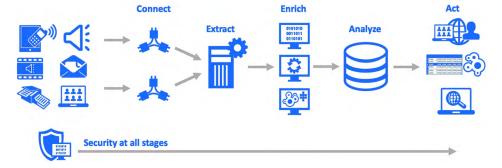
OpenText™ IDOL Connectors synchronize data from over 150 types of repositories, including:

- Local and network filesystems
- Internet servers, websites, and social media feeds
- Cloud-based storage such as Microsoft OneDrive, Amazon S3, Google Drive, and Azure Blob
- Email servers such as Microsoft Exchange (on-prem and online), IMAP, and POP3
- Database servers such as SQL Server, Oracle, and MySQL
- Document Management and Collaboration systems such as Microsoft SharePoint, Teams, Confluence, Lithium, and Alfresco

Select connectors can operate in a bidirectional mode, allowing you to take actions directly on the source repository such as hold or delete, quickly reducing the risk associated with manage-in-place.

Decryption

Keep up with emerging data storage trends by inspecting the content of files that have been protected by Microsoft's Azure Rights Management (RMS), part of AIP and MIP. IDOL

Panopticon can be used with new or existing workflows to allow unrestrained access to the original unencrypted content of protected files with little to no fuss.

File Type Detection

Reduce the risk of misprocessing crucial information or wasting valuable CPU time on irrelevant files. Instead of relying on falsifiable file name extensions or short magic numbers, OpenText™ IDOL KeyView forensically examines each file, focusing on the most differentiating characteristics first and going as deep as needed to resolve ambiguity, resulting in faster answers and a lower error rate.

IDOL KeyView goes beyond MIME type, clearly identifying files with non-existent or ambiguous MIME types (e.g., application/octet-stream), adding detail such as character set identification, encryption status, format classification, and format version, giving you precision on which to base your downstream routing and processing decisions.

Content Extraction

Don't miss a thing, with the richest text extraction (filtering) technology available: IDOL KeyView. Get deep visibility into a wide variety of document formats, extracting content

including body text, visible components (such as headers, footers, and section names), invisible text (such as notes, tracked changes, accessibility layers, and placeholder text), embedded, cached, and encapsulated objects, as well as file metadata (such as format specific fields, XMP, XrML, and classification).

IDOL KeyView supports pipelining within documents, so time-critical downstream processing can start earlier. This results in lower latency and enables you to meet tight SLAs.

OCR

Extend visibility into image-based sources using Optical Character Recognition (OCR) on a wide variety of noisy inputs, including:

- Scanned documents and screenshots
- Subtitled television footage, such as news channels with banners
- Photographs of real-world scenes, such as street signs
- Photographs of documents, such as from a smartphone

When using IDOL KeyView, OCR of scanned documents can be enabled with a single configuration change for ultimate ease-of-deployment, adding visibility into digitally archived content with no code changes.

Document Preview

Improve usability by providing high-fidelity viewing (HTML5) and printing (PDF) capabilities for numerous document types including word processing, spreadsheet, presentation, graphic, multimedia, and archive formats, all without the need to install native applications for each format.

Entity Extraction and Redaction

Meet compliance and governance requirements such as GDPR, CCPA or corporate policies by using the IDOL Eduction engine to find sensitive data and other entities of interest in data being processed, within strict timelines.

OpenText™ maintains an extensive library of highly optimized grammars ranging across languages, countries, and industries allows for easy identification of entities including PII, PHI, PCI and many other regulated areas. Finding and securing documents containing these entities reduces compliance risk and saves labor hours.

Redact sensitive content in text, images, and audio (including within video) to comply with legal and corporate mandates.

Sentiment Analysis

Improve customer satisfaction by discovering sentiment trends on existing or previously unknown entities. IDOL operates at the clause level when faced with challenging constructs such as "the food was terrible but the front-of-house staff were great", giving better precision and more accurate results on which to base decisions.

Rich Media Analysis

Understand immediately the meaning within rich media assets by distilling facts from video, image, and audio content in real time.

Video and Image Analysis

- Face Detection, Recognition, and Analysis
- Barcode Recognition
- Object Recognition
- Color Analysis
- Automatic Number Plate Recognition (ANPR)
- Scene Analysis
- People Counting and Tracking

Audio Analysis

- Speech-to-Text
- Phonetic Search
- Speaker Identification
- Spoken Language Identification
- Audio Matching
- Transcript Adherence
- Transcript Alignment

Conceptual Classification

Improve user efficiency by automatically assigning documents to categories according to what concepts the document and the category have in common. Train categories conceptually, either with a set of representative documents or by writing free text, or by writing Boolean expressions that describe the relevant concepts.

Conceptual categories provide an intuitive way to navigate documents, and to filter out unimportant information from a high-volume set of information, allowing focus on what matters most, saving time and money.

Cross-Document Analytics

Create understanding of large document sets by automatically identifying common information across documents.

Clustering. Find clusters of related concepts or documents, either across an entire document set, or within a subset such as a snapshot or a dynamic query result set. Automatically create and populate a taxonomy of your documents, giving cognitive structure to previously dark data.

What's Hot/What's New. Surface areas of interest that have become particularly important, or that have not been seen before, using time-based trend detection of topics, to reduce the risk of getting out of date in an ever-changing data landscape.

Structured Analytics. Analyze the structured fields of documents to identify and tag patterns and trends, including proximate analysis for 'near' matches to find answers in an imperfect world.

Community. Increase productivity by understanding who knows what in your user base. Find subject matter experts, interested parties, and thought leaders, providing more efficient collaboration and saving time.

Search. Speed workflow by finding information quickly and accurately, using IDOL's highly configurable core information retrieval engine. Combine concepts, facets, geolocation, personalization, fuzzy logic and more to give your users the best possible search experience.

Virtual Assistant

Increase customer satisfaction and user efficiency while lowering staffing costs by providing a real time, interactive conversation using a natural, human-friendly textual interface to answer natural language questions and automate common processes without employing additional staff.

Better Together

Each IDOL component is designed to be embedded in a larger product or service and many components offer additional benefits when multiple IDOL products are used together.

For example, by pipelining content from file filtering into PII detection, you can drastically improve the first response time during risk analysis, meaning end users are not kept waiting for buffers to fill and can access their documents without delay.

Get the benefits of ecosystem-aware components and see how IDOL can accelerate your offering.

Learn more at
www.opentext.com