OpenText Intelligent Classification

A fast, powerful, flexible way to unlock the value hidden in unstructured text by extracting terms, concepts, entities, sentiments, emotions and more to yield business insights

Benefits

- Find new value in content and derive more insight with semantic metadata
- Improve customer support by connecting more efficiently to desired information
- Boost insight and responsiveness by analyzing social media and news content
- Streamline information governance and regulatory compliance by indexing and enriching stores of content

Organizations can now quickly and cheaply access billions of pages of text from internal documents, emails, social media, web content and more. But, without the right tools, they often fail to tap into its insights. The potential lost value is enormous, since nearly **80 percent of all enterprise information** is unstructured data.[1]

The OpenText™ Intelligent Classification application gives organizations visibility into unstructured, text-based content by using semantic technologies and metadata to improve its findability. This powerful natural language processing solution can collect, sift, correlate and assign meaning to text from emails, PDFs, web content, social media and other internal and external data sources. This includes identifying key people, topics, dates, events, themes and even the mood and subjectivity of the document. By some estimates, knowledge workers spend one-sixth of their time hunting for information and 44 percent of those searches come up empty-handed.[2] With OpenText Intelligent Classification, knowledge workers can use this output in an analytic context to more easily spot patterns and trends to gain new insights, make better decisions and free up time.

These capabilities can make information governance more accurate and efficient, cut down on redundant, obsolete or trivial information and extend the lifespan of useful data. OpenText Intelligent Classification automates practices and enables previously unheard-of performance boosts to processes, such as measuring customer or citizen sentiment, legal discovery, risk assessment, regulatory compliance, customer service and product personalization.

1   TechRepublic, *Unstructured data: A cheat sheet.* (2017)
2   IDC, *The Knowledge Quotient: Unlocking the Hidden Value of Information Using Search and Content Analytics.* (2014)

Drawing on powerful machine learning technology, the solution comes with built-in libraries of terms (including many important business concepts), which allow organizations to make the most of scarce data science talent. Further, it can be "taught" nearly any other topic, in multiple languages.

Find new value in content and derive more insight

Textual content inherently contains value, but it is only valuable if the right information can be found by the right people, quickly. OpenText Intelligent Classification goes beyond simple tagging to extract what the content is about, linguistically parsing every sentence to create rich metadata that makes it easier to find and add meaning to key names, places, dates, concepts and much more. Organizations find it easier to track down, summarize, analyze and spot patterns or trends in the petabytes of information they gather, such as customer comments, patient records or legally required financial disclosures, and create user-friendly reports or blend with content from other sources to create even more high-value assets.
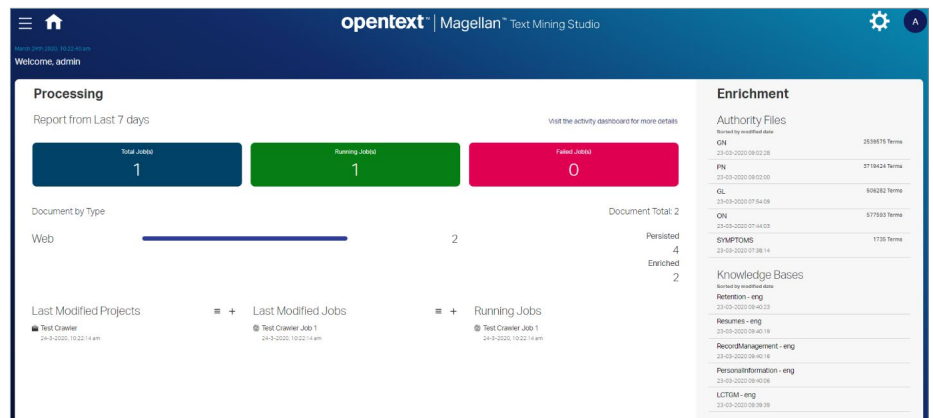
With the OpenText Intelligent Classification Studio, it is easy to edit and add new categories of knowledge.

Improve customer support

OpenText Intelligent Classification efficiently serves up the right pieces of text from among millions or billions of pages to complete an order record, address a service question or personalize a product recommendation.

Boost insight and responsiveness to public opinion

Customer or citizen sentiment and emotion can be gauged by analyzing the meaning and mood of social media and news content and spotting trending terms. OpenText Intelligent Classification can identify whether the prevailing sentiment of the text is positive, neutral or negative, whether it reflects subjective opinion, and which emotions (anger, anticipation, disgust, fear, joy, sadness, surprise, and trust) are expressed and their classifications.

Streamline information governance and regulatory compliance

Information governance and regulatory compliance become more reliable and efficient with OpenText Intelligent Classification, which can intelligently index and enrich content archives to help users spot content that needs to be retained, archived or shared and clear out the clutter.

Enable fast deployment to the platform of choice

OpenText Intelligent Classification is easy to set up and operate, with convenient pre-configured vocabularies and taxonomies, plus flexible machine learning options to digest new topics. It comes equipped with a range of crawlers and connectors to a wide range of popular enterprise content repositories (including file shares), a management console and a visually appealing, self-service UI. A complete list of connectors included with OpenText Intelligent Classification appears on page 5.

OpenText Intelligent Classification also allows organizations to deploy on the hosting solution of their choice, whether on-premise, through four-node Windows or Linux deployments, hybrid, or in the cloud. The OpenText Intelligent Classification cloud-native edition provides a modern, containerized architecture, supported by Docker and Kubernetes, that delivers high levels of performance, flexibility, and scalability, and can be deployed to the OpenText Cloud or other platforms of choice.

Seamless EIM integrations

OpenText Intelligent Classification readily integrates with OpenText's other award winning EIM applications and acts as a digital accelerant to create quick, effective solutions for a range of business uses including AI-enhanced content capture, predictive asset performance optimization, and intelligent shopping recommendations. For example, the OpenText™ Webs' content management system uses OpenText Intelligent Classification for automatic content tagging, summarization and taxonomy creation, promoting easier, more accurate content access, targeted personalization and AI-driven content suggestions.

How it works

OpenText Intelligent Classification uses a best-of-breed approach that combines statistical methods with machine learning and knowledge engineering to analyze each piece of content and assign it a linguistic "fingerprint" through metadata. It also processes text, extracts entities (places, people, organizations and anything else with a name), concepts and the sentiment and emotion of each document. It summarizes documents by identifying key sentences, then profiles and categorizes them. To see it in action, visit http://magellan-text-mining.opentext.com and paste in the text of your choice.
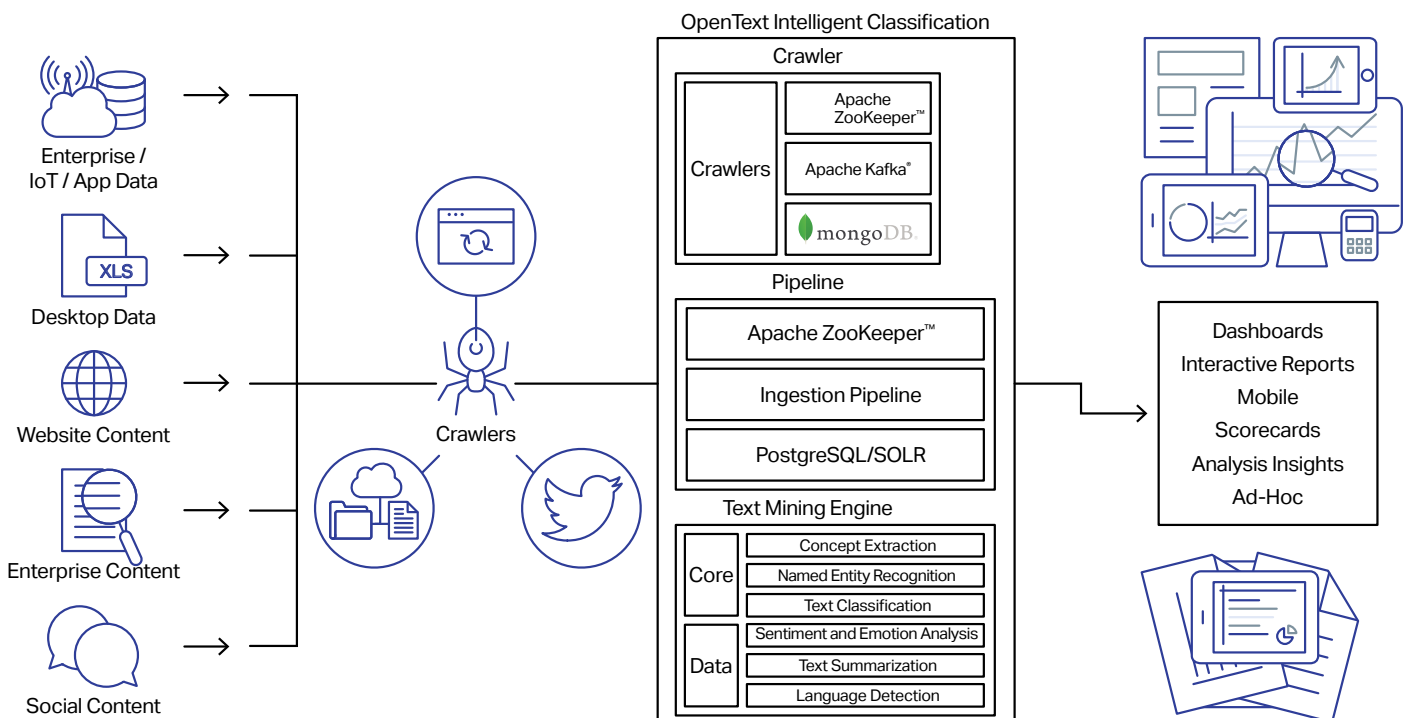
OpenText Intelligent Classification processing

Concept extraction	Identifies meaningful terms and noun and verb phrases from documents, using algorithms to extract core concepts
Named entity recognition (NER)	Locates and extracts places, people, organizations and anything else with a name
Text classification	Indexes and sorts documents by classification and identifies relevancies Analyzes concepts and queries them against extensible models and knowledge bases to create document profiles according to a defined taxonomy
Text summarization	Identifies key sentences in a document and uses them to create a summary, liberating staff from an otherwise manual chore
Sentiment and emotion analysis	Detects whether a document or even a sentence expresses a negative, positive, or neutral connotation, is objective or opinion-based, and to what degree it expresses anger, anticipation, disgust, fear, joy sadness, surprise and trust
Language Detection	Finds the language of text or paragraphs within text.

OpenText Intelligent Classification languages

OpenText Intelligent Classification supports these languages with dedicated natural language processing and one or more available annotators out of the box:	Arabic, Chinese, Dutch, English, French, German, Hebrew, Italian, Japanese, Portuguese and Spanish
OpenText Intelligent Classification supports basic concept and named entity extraction in 25 additional languages:	Bulgarian, Catalan, Croatian, Czech, Danish, Estonian, Finnish, Greek, Hungarian, Irish, Icelandic, Latvian, Lithuanian, Norwegian, Persian, Polish, Romanian, Russian, Slovak, Slovene, Swedish, Turkish, Ukrainian and Vietnamese

OpenText Intelligent Classification architecture, from data ingestion (L) to actionable output (R)

OpenText Intelligent Classification at work

A major food and beverage company knew its success depended in large part on public opinion. Consumers' changing tastes guide the choice of products to launch, expand and discontinue. Levels of brand awareness influence marketing campaigns, while complaints might indicate problems the company could fix, such as quality or shipping times. However, traditional opinion-gathering methods, such as surveys, focus groups and media audits, are slow and expensive.

The company wanted to monitor social media and customer feedback, not just tally hashtags but report on the concepts and emotions people have about the company.

OpenText Intelligent Classification offered a fast, accurate, cost-effective way to view and analyze discussions of the company and its products, not only the terms and attitudes it knew to track but some it didn't expect. The company can now target its marketing, product development and consumer outreach efforts more accurately, improving customer satisfaction and its own profitability.

Connectors	
Included	
OpenText	Content Server 10.x (LAPI*, CWS*)
	Content Server 16 (CWS, REST)
	Content Server 16.2-16.5 (CWS, REST)
	Archive Center (versions 10.5 and 16)
	OpenText™ Legal Content Management (versions 5.3.1, 10, and 16.2-16.5)
	OpenText™ Digital Asset Management (versions 7.3.1 and 16.x [via REST v4])
	OpenText™ Information Archive
	Tempo Box
	OpenText™ Documentum™ Content Management (versions 6.6, 7.1, and 16.x)
Microsoft	Windows fileshare
Add-Ons	
IBM	FileNet P8 4.5.1
	FileNet P8 5.2
Cloud storage	Box.com
	Dropbox
	Google Gmail
	Google Drive
Microsoft	Microsoft Exchange 2010
	Microsoft SharePoint 2010
	Microsoft SharePoint 2013
	Microsoft SharePoint 2016
	Microsoft SharePoint Online
Generic	CMIS

OpenText Intelligent Classification offers crawlers that can analyze documents by connecting with a wide range of popular content management tools.

* Note: CWS stands for "Content Web Service" interface. LAPI stands for "Livelinks API."

OpenText Intelligent Classification services	
Modeling services	Provides a growing collection of REST APIs that extract and enrich unstructured content with semantic metadata and support modeling. Users can create and manage custom taxonomy and authority files, then train and benchmark classification models and/or rules for them and use the modeling services to customize the provided semantic annotations based on business requirements.
Crawling services	Enables users to access, extract and process content from supported content repositories. These crawlers include native and direct connectors into OpenText EIM solutions, such as OpenText™ Content Server, OpenText™ Documentum™ Content Management, OpenText™ Media Manager, OpenText™ Archive Center, OpenText™ Information Archive, OpenText™ Tempo Box, OpenText™ Legal Content Management and a generic CMIS connector. Crawlers that reach directly into popular third-party tools, such as Box, Dropbox, Gmail™, Google Drive™, IBM® FileNet®, Microsoft® Exchange and SharePoint®, are readily available as add-ons.
Annotation service	Powers the core capabilities of OpenText Intelligent Classification by leveraging classification trees within the models. This service enables content enrichment with categories/classification, named entities, sentiment and emotion analysis, and key concepts at the document, sentence or entity level.
OpenText Intelligent Classification Studio	Offers a convenient, visually appealing user interface to control the content analysis process.
Persistence layers	Includes RDBMS, Hive, Apache Solr and Apache Spark services.

opentext™