

# OpenText Intelligent Classification

Leverage powerful natural language processing to analyze unstructured content and unlock the insights held within

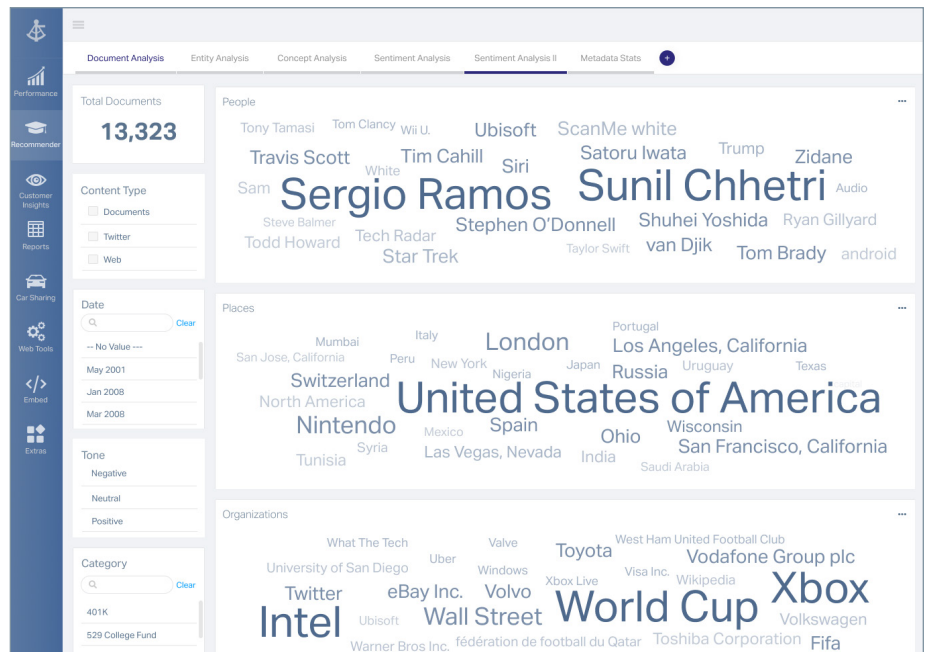


## Benefits

- Reduce manual effort with automated document review and capture of semantic metadata
- Improve the content discovery process with bespoke indexing and classification
- Extract meaningful data and insights from complex unstructured sources quickly and consistently
- Built using out-of-the-box machine learning capabilities through pre-trained and custom models for text categorization and text mining

OpenText Intelligent Classification enables enterprises with an all-in-one solution that combines data and entity extraction, text analytics, natural language processing, and semantic analytics to garner powerful knowledge from their unstructured content, and leverage it to better manage and grow their businesses. Unstructured content, such as articles, documents, abstracts, emails, reports, captions, customer chat logs, and social media content, store invaluable data that is typically inaccessible to the organization or requires tedious and time-consuming manual analysis.

OpenText Intelligent Classification empowers organizations to extract meaningful information from each textual asset, and automatically tag and enrich it with semantic metadata. This deepens an organization's ability to gather insights and improves the findability of content by better connecting individuals to more relevant pieces of content. The platform provides robust metadata relationship retrieval and building capabilities, and the semantic metadata provided by OpenText Intelligent Classification can be blended with other metadata types (e.g. system, editorial). This capability also extends to Enterprise Content Management systems and analytical resources, such as OpenText Intelligent Classification dashboards, to enhance business insights processes and deliver an extensive content intelligence and analysis experience.



OpenText Intelligent Classification insights can be displayed in an easy-to-use dashboard when used with the OpenText Intelligence.

## OpenText Intelligent Classification modules

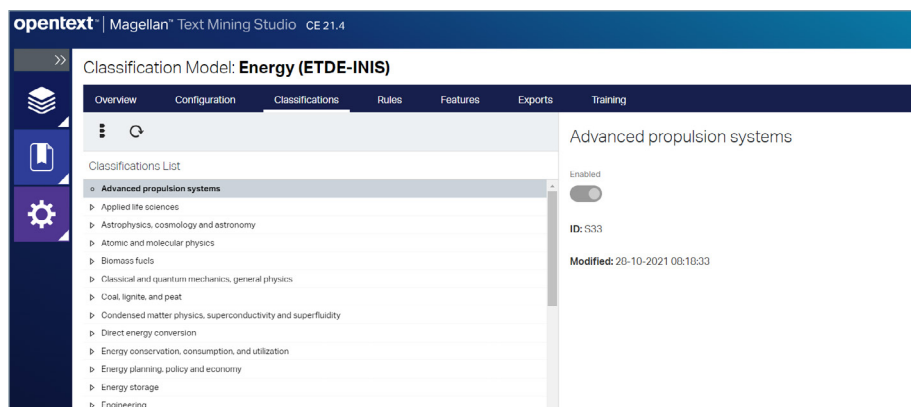
OpenText Intelligent Classification is comprised of six modules that provide an end-to-end process to extract, analyze, and process specific types of semantic metadata:

Feature	Description
<b>Concept extraction</b>	<p>Identifies meaningful keywords and key phrases, extracts core concepts following grammatical patterns of a noun, verb phrase or statistical tokens, and can be configured to specifications for unique use cases such as large or short texts</p> <p>Concept extraction operates in four steps:</p> <ol style="list-style-type: none"> <li>1. Text parsing and word cropping: Analyze and distinguish keyword candidates from abbreviations, delimiters, etc.</li> <li>2. Part of speech tagging: Assign a grammatical role to each keyword</li> <li>3. Extraction of part of speech patterns: Identify and expose relevant grammatical patterns such as simple nouns and complex noun phrases</li> <li>4. Expose metadata: Concepts are exposed along with the information regarding their frequency, relevancy, and position.</li> </ol>

## Out-of-the-box Categorizer Knowledge Bases

- Business and Finance
- General Business
- International Press Telecommunications Council (IPTC)
- Industry Classification Benchmark (ICB)
- Library of Congress Thesaurus for Graphic Materials (LCTGM)
- Energy Technology Data Exchange and International Nuclear Information System (ETDE-INIS)
- Generally Accepted Accounting Principles for Commerce and Industry (GAAP-CI)
- Records Management
- Resumes
- Retention
- Personal Information

Feature	Description
<b>Named entity recognition (NER) &amp; information extraction</b>	Locates and identifies terms or sequences of terms that can be referred to with a name, such as organizations, geographic locations, trademarks, events, people names, and diseases. For example, "Open Text Corporation" would be an organization entity, "North America" a geographic entity and "Albert Einstein" a person name entity. Through advanced machine learning and authority files, OpenText Intelligent Classification can also recognize and associate different names for the same entity to improve content findability. OpenText's Named Entity Extractor includes normalization algorithms to ensure standardized entity assignments across all content. Named entities are exposed along with relevancy and confidence scores, sub-term/parent positions, and their length. These entities can be also exposed in a context showing subjectivity and tonality information around each entity sub-term.
<b>Text classification</b>	Indexes and sorts documents by classification along with the classification weight and relevancy ranking. Classifications are topics inferred from the body of a document; they are identified and associated with documents even if the category name is not explicitly stated in the text.
<b>Text summarization</b>	Identifies key sentences in a document and uses them to create a brief overview according to topics of interest.
<b>Sentiment and emotion analysis - Subjectivity, tone, and emotion</b>	Determines the subjectivity (fact or opinion), tone (positive, negative, or neutral), and emotions (anger, anticipation, distrust, fear, joy, sadness, surprise, and trust) of the document. Sentiment analysis is performed first on a sentence level, then on the document level and around named entities. The Sentiment Analyzer module can also perform emotion analysis to determine the emotion (anticipation, anger, disgust, fear, joy, sadness, surprise, trust) within text, along with their variations depending on the weight shown for each.
<b>Language detection</b>	Recognizes the language(s) used throughout the document and automatically categorizes the document accordingly.



With the OpenText Intelligent Classification Studio it's easy to manage and edit existing knowledge bases, or import new ones.

## ISO/NISO Standard Authority Files

- Organization names
- Person names
- Geopolitical locations
- Geophysical names
- Trademark
- Events
- Life sciences: Diseases, Symptoms, Drugs, Physiology
- Features
- Currency
- Date and time
- Computer and internet
- Twitter
- Personally Identifiable Information (PII): SSN, Passport, Phone Number, Bank Account, Address, Driver's License

## Connectivity

Straightforward connections that enable your organization to extract, tag, and analyze unstructured content from a vast range of sources, all managed, controlled, and operationalized from a single location.

Feature	Description
<b>OpenText</b>	<ul style="list-style-type: none"><li>• Archive Center</li><li>• Content Server</li><li>• Documentum™ Content Management</li><li>• eDOCS Content Management</li><li>• Information Archive</li><li>• OpenText™ Digital Asset Management</li><li>• OpenText Tempo Box</li></ul>
<b>Connectors for third-party repositories</b>	<ul style="list-style-type: none"><li>• Box.com</li><li>• CMIS</li><li>• Dropbox</li><li>• Google Drive</li><li>• Google Gmail</li><li>• IBM FileNet</li><li>• Microsoft Exchange</li><li>• Microsoft SharePoint</li></ul>
<b>Supported databases</b>	<p><b>OpenText Intelligent Classification Conversion Service</b></p> <ul style="list-style-type: none"><li>• Microsoft SQL Server</li></ul> <p><b>OpenText™ Intelligent Classification Pipeline</b></p> <ul style="list-style-type: none"><li>• Apache Hive</li><li>• Apache Spark</li><li>• Microsoft SQL Server</li><li>• MySQL Enterprise</li><li>• PostgreSQL</li></ul> <p><b>ECM Crawling Service</b></p> <ul style="list-style-type: none"><li>• Microsoft SQL Server</li><li>• Oracle</li><li>• PostgreSQL</li></ul> <p><b>OpenText Integration Services</b></p> <ul style="list-style-type: none"><li>• Microsoft SQL Server</li><li>• Oracle</li><li>• PostgreSQL</li></ul> <p><b>ECM repositories</b></p> <ul style="list-style-type: none"><li>• OpenText Content Server</li><li>• OpenText Tempo Box</li><li>• Windows file shares</li></ul>

## Methodology and analysis

OpenText Intelligent Classification uses a variety of carefully chosen methodologies that complement each other, and can be tailored to your specific vocabularies, so that it can identify and extract the most relevant semantic metadata quickly and accurately. It also draws upon powerful machine learning capabilities, comes with built-in libraries of terms (including many important business concepts), pre-trained text categorization models, and can be trained on nearly any other topic across a multitude of languages.

Feature	Description
<b>Pattern matching</b>	<ul style="list-style-type: none"><li>• Statistical - Translates segments of text into paragraphs, sentences, tokens, and n-grams, along with attributes for each segment</li><li>• Linguistic – Identification of words and phrases that are likely to be concepts and patterns and transforms tokens and n-grams from “sequences of items” to real lexical units</li></ul>
<b>Machine learning</b>	<ul style="list-style-type: none"><li>• Leverage powerful families of algorithms, including Naïve Bayes, Boosting Algorithm, Keywords computing, Decision trees, to train ML models and improve prediction accuracy.</li><li>• Following OpenText’s metadata assignment process, users can apply post-processing algorithms to review the metadata based upon the factors most important to them</li></ul>
<b>Knowledge engineering</b>	<ul style="list-style-type: none"><li>• The knowledge engineering functionalities embedded in OpenText Intelligent Classification empower users to create and manage their own taxonomies and authority files for their own business-specific use cases.</li><li>• Multiple Intelligent Classification modules use knowledge engineering to improve ML prediction accuracy through creation of a training set and model, optimize the model based on statistical observations, and configure it to specific customer needs.</li></ul>
<b>Supported languages</b>	<p><b>OpenText Intelligent Classification supports these languages with dedicated natural language processing and one or more annotators available out of the box:</b></p> <ul style="list-style-type: none"><li>• Arabic, Chinese, Dutch, English, French, German, Hebrew, Italian, Japanese, Portuguese, Spanish</li></ul> <p><b>OpenText Intelligent Classification supports basic concept and entity extraction in 25 additional languages</b></p> <p>Dedicated Tokenizer</p> <ul style="list-style-type: none"><li>• Czech, Persian, Polish, Russian, Turkish</li></ul> <p>Multilingual Tokenizer</p> <ul style="list-style-type: none"><li>• Bulgarian, Catalan, Croatian, Danish, Estonian, Finnish, Greek, Hungarian, Icelandic, Irish, Latvian, Lithuanian, Norwegian, Romanian, Slovak, Slovenian, Swedish, Ukrainian, Vietnamese</li></ul>

## Resources

[Try OpenText Intelligent Classification ›](#)

[Keep up to date ›](#)

[Read the latest blog posts ›](#)

## Methodology and analysis (Cont'd)

Feature	Description
<b>Text mining – Simple mapping parameters keys</b>	<ul style="list-style-type: none"><li>• Summary</li><li>• Subjectivity score and its distribution</li><li>• Tone type (positive/negative/neutral)</li><li>• Positive/Negative/Neutral tone scores and their distribution</li><li>• Simple concept</li><li>• Complex concept</li><li>• Named entities</li><li>• Classifications</li><li>• Emotions (anger, anticipation, disgust, fear, joy, sadness, surprise, trust) and their weight</li></ul>

## Text mining services

Feature	Description
<b>Modeling &amp; benchmarking services</b>	Provides a growing collection of REST APIs that extract and enrich unstructured content with semantic metadata and support text classification modeling. Users can create and manage custom taxonomies 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.
<b>Crawling services</b>	Enables users to access, extract and process content from multiple supported content repositories including enterprise content management, web, and X.
<b>Annotation service</b>	Powers the core capabilities of OpenText Intelligent Classification by leveraging taxonomies and authority files. This service enables summarization, and content enrichment with categories/classifications, named entities, key concepts, emotion and sentiment analysis at the document, sentence or entity level.
<b>OpenText Intelligent Classification Studio</b>	Offers a convenient, visually appealing user interface to control the content analysis process, from content access through semantic enrichment and metadata federation.
<b>Persistence layers</b>	Includes RDBMS, Hive, Apache Solr and Apache Spark services.