opentext™

DATA SHEET

Magellan Text Mining

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

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 Magellan Text Mining 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.

Magellan Text Mining 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 Magellan Text Mining 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 Magellan dashboards, to enhance business insights processes and deliver an extensive content intelligence and analysis experience.
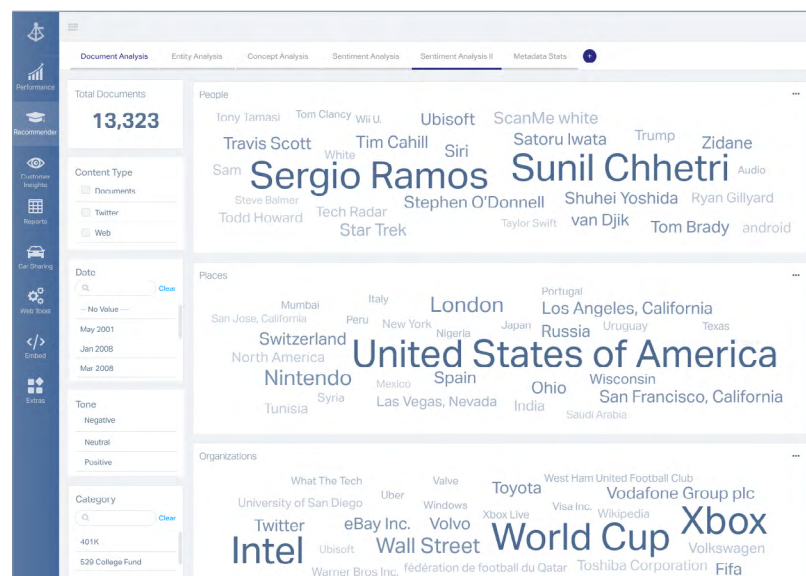
Magellan Text Mining insights can be displayed in an easy-to-use dashboard when used with the Magellan BI & Reporting.

opentext™

Associated OpenText products

- Magellan Content Analytics Suite
- Magellan Platform
- Magellan AI-powered augmented Capture
- Magellan AI-augmented Voice of the Customer
- Magellan AI-driven Content Advisor

Magellan Text Mining modules

Magellan Text Mining is comprised of six modules that provide an end-to-end process to extract, analyze, and process specific types of semantic metadata:

| Feature	Description
Concept extraction	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 Concept extraction operates in four steps: 1. Text parsing and word cropping: Analyze and distinguish keyword candidates from abbreviations, delimiters, etc. 2. Part of speech tagging: Assign a grammatical role to each keyword 3. Extraction of part of speech patterns: Identify and expose relevant grammatical patterns such as simple nouns and complex noun phrases 4. Expose metadata: Concepts are exposed along with the information regarding their frequency, relevancy and position.
Named entity recognition (NER) & information extraction	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, Magellan Text Mining can also recognize and associate different names for the same entity to improve content findability. Magellan'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.
Text classification	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.
Text summarization	Identifies key sentences in a document and uses them to create a brief overview according to topics of interest.

opentext™

- 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

Magellan Text Mining modules (Cont'd)

Feature	Description
Sentiment and emotion analysis - Subjectivity, tone, and emotion	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.
Language detection	Recognizes the language(s) used throughout the document and automatically categorizes the document accordingly.
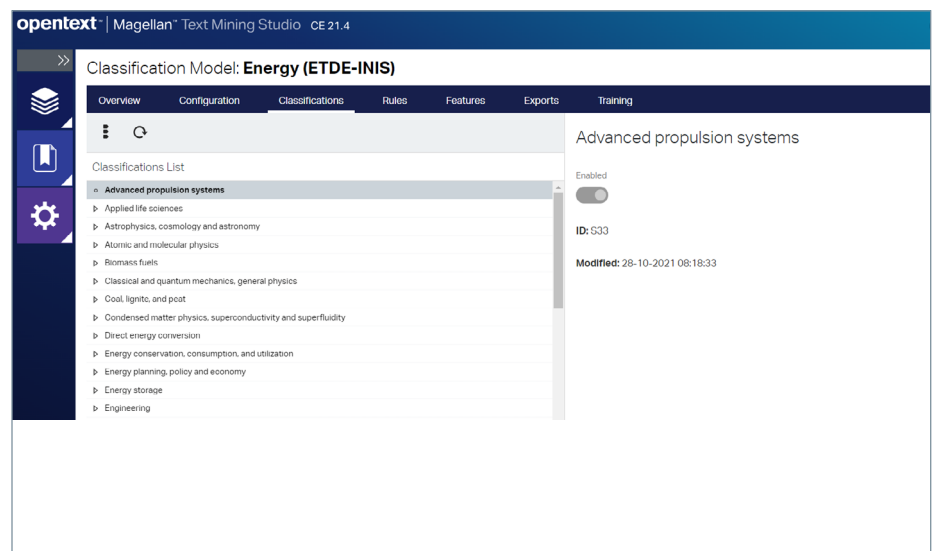
With the Magellan Text Mining Studio it's easy to manage and edit existing knowledge bases, or import new ones.

opentext™

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
OpenText	• Archive Center • Content Server • Documentum • eDocs • InfoArchive • Media Manager • Tempo Box
Connectors for 3rd party repositories	• Box.com • CMIS • Dropbox • Google Drive • Google Gmail • IBM FileNet • Microsoft Exchange • Microsoft SharePoint
Supported databases	**Magellan Text Mining Conversion Service** • Microsoft SQL Server **Text Mining Pipeline** • Apache Hive • Apache Spark • Microsoft SQL Server • MySQL Enterprise • PostgreSQL **ECM Crawling Service** • Microsoft SQL Server • Oracle • PostgreSQL **Magellan Integration Services** • Microsoft SQL Server • Oracle • PostgreSQL **ECM Repositories** • OpenText Content Server • OpenText Tempo Box • Windows file shares

opentext™

Methodology & Analysis

Magellan Text Mining 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
Pattern matching	• Statistical - Translates segments of text into paragraphs, sentences, tokens, and n-grams, along with attributes for each segment • 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
Machine learning	• Leverage powerful families of algorithms, including Naïve Bayes, Boosting Algorithm, Keywords computing, Decision trees, to train ML models and improve prediction accuracy. • Following Magellan's metadata assignment process, users can apply post-processing algorithms to review the metadata based upon the factors most important to them
Knowledge engineering	• The knowledge engineering functionalities embedded in Magellan Text Mining empower users to create and manage their own taxonomies and authority files for their own business-specific use cases. • Multiple Text Mining 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.
Supported languages	**Magellan Text Mining supports these languages with dedicated natural language processing and one or more annotators available out of the box:** • Arabic, Chinese, Dutch, English, French, German, Hebrew, Italian, Japanese, Portuguese, Spanish **Magellan Text Mining supports basic concept and entity extraction in 25 additional languages** Dedicated Tokenizer • Czech, Persian, Polish, Russian, Turkish Multilingual Tokenizer • Bulgarian, Catalan, Croatian, Danish, Estonian, Finnish, Greek, Hungarian, Icelandic, Irish, Latvian, Lithuanian, Norwegian, Romanian, Slovak, Slovenian, Swedish, Ukrainian, Vietnamese

opentext™

Methodology & Analysis (Cont'd)

Feature	Description
Text mining – Simple mapping parameters keys	• Summary • Subjectivity score and its distribution • Tone type (positive/negative/neutral) • Positive/Negative/Neutral tone scores and their distribution • Simple concept • Complex concept • Named entities • Classifications • Emotions (anger, anticipation, disgust, fear, joy, sadness, surprise, trust) and their weight

Text Mining Services

Feature	Description
Modeling & bench-marking services	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
Crawling services	Enables users to access, extract and process content from multiple supported content repositories including ECM, Web, and Twitter.
Annotation service	Powers the core capabilities of Magellan Text Mining 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.
Magellan Text Mining Studio	Offers a convenient, visually appealing user interface to control the content analysis process, from content access through semantic enrichment and metadata federation.
Persistence layers	Includes RDBMS, Hive, Apache Solr and Apache Spark services.

opentext™

⚙ Try Magellan Text Mining

🐦 Keep up to date

💬 Read the latest blog posts

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**.

Connect with us:

- **OpenText CEO Mark Barrenechea's blog**
- **Twitter** │ **LinkedIn**

opentext.com/contact