



SAS® Text Analytics

Maximize the value buried in your unstructured content



Overview

Confronted with big data issues arising from the deluge of unstructured content, many organizations struggle to get the best possible value from text data. Because of the ambiguity and numerous ways that text represents similar concepts, information in text-based data is not easy to discern, quantify, analyze or exploit. And many organizations cannot combine text-based information with structured data – so it's impossible to get a full, accurate view of the enterprise.

SAS Text Analytics turns text data into an organizational asset by helping you automatically assess, analyze and act upon the insight buried in electronic text – including social media content, call center logs, survey data, emails, loan applications, service notes, and insurance or warranty claims. You can incorporate the new information from SAS Text Analytics as metadata to improve search and relevancy of retrieved documents, as well as predictive accuracy. As a result of having reliable, comprehensive and documented information, you can make more effective, proactive business decisions, streamline priorities and produce more meaningful reports.

Challenges

- **Maximizing the business value of unstructured data.** With the volume, velocity and complexity of big data, businesses struggle to automatically integrate and obtain contextual insight from all types, sources and locations of content.
- **Understanding sentiment expressed in social media.** Analyzing information from the Web can be labor-intensive, but businesses increasingly rely on these sources to discover trends, opportunities and risks related to their products and services.
- **Ineffective search and retrieval.** Queries often return large collections of irrelevant materials that require manual review to determine which one should be used.
- **Difficulty analyzing text-based data.** Because of the ambiguity and numerous ways to represent concepts, text-based data is not easy to discern, quantify or analyze.



Extract information and create knowledge from text-based documents

■ Key Benefits

- **Eliminate content chaos** in your organization by retaining and distributing only relevant, accurate and up-to-date information.
- **Discover explicit relationships and trends** in unstructured and semistructured data for more informed and accurate decision making.
- **Analyze information** from the Web, social networking sites and internal electronic document collections to understand expressed sentiments and their effect on the bottom line.
- **Stay current** on trends, research and competitive positioning in your marketplace by automatically evaluating internal and external materials.
- **Be proactive** through early identification of product or service issues, before they affect sales and customer satisfaction.
- **Accurately categorize content** to enable efficient information organization, access, distribution and findability.
- **Gain a more complete view of organizational issues** by creating a structure for contextual analysis and integration with structured data.

How SAS® Can Help

Streamline Business Processing for Large Volumes of Multilingual Content

SAS Text Analytics uses natural language processing (NLP) and advanced linguistic techniques to automatically categorize and analyze large volumes of content for entities, facts and events. Analysis of the text creates metadata, documented taxonomies, and linguistic rules and concept definitions that can be automatically applied to large document collections for fast, accurate organization – and used to trigger real-time business processes.

Discover and Extract New Knowledge from Text Documents

With a rich suite of tools, your text data is transformed into a usable, systematic format. This helps to facilitate document classification, discover explicit associations between documents and cluster documents into categories for new insights across collections.

Distill Documents with Built-In Subject-Matter Expertise

Advanced linguistic and parsing technologies can standardize subject-matter expertise within your organization. When defined to the system, documents are automatically processed to these knowledge-driven rules and concepts. Then, documents can be distributed in real time, distilled into existing workflows and updated to transaction systems – automating information sharing and content distribution.

Analyze Social Media and Capture Relevant Product Reviews

Automated processes locate and analyze digital content in real time to help you discover and interpret the emotions behind the qualitative information posted on the Internet or from internal communications. Relying on these processes allows your marketing, brand and logistics management professionals to overcome the limitations of manually discovering, interpreting and annotating sentiments using search tools or doing time-consuming manual reviews of text.

Link Document Repositories with Consistently Defined Relationships

By associating disparate collections with systematic definitions and rules, you can semantically relate existing text repositories and assets. This maximizes the value of your organization's text collections by creating a defined lens to the information stores and bringing them to the forefront when related information is examined. It also improves information retrieval activities and automates the delivery of related topics.

Integrate Text-Based Information with Structured Data

This solution enables you to consolidate structured data with unstructured data (free-form text) to provide complete views and meaningful insights within an integrated predictive modeling environment. Analysis performed using both types of data will produce a more accurate and complete view of your organization.

Components

SAS® Enterprise Content Categorization

Managing enterprise content effectively and efficiently, and as a strategic asset, requires a common, underlying informational structure. SAS Enterprise Content Categorization applies natural language processing (NLP) and advanced linguistic techniques to identify key topics and phrases in electronic text so you can automatically categorize large volumes of multilingual content that is acquired, generated or exists in a repository. It correctly parses, analyzes and extracts content for entities, facts and events to create metadata tags that index documents in a collaborative taxonomy management environment. As a result, people across your organization can quickly find the relevant content they need, when they need it, at the level of granularity required. The software uses sophisticated linguistic rules to consistently organize, index and trigger dependent informational activities in real time. This drives faster, more efficient information organization, access, findability and knowledge sharing – and reduces overhead associated with processes such as manual tagging and retrospective indexing.

Add-Ons for SAS® Enterprise Content Categorization

Extend your organization's existing content categorization capabilities and customize your content management solution to improve document-centric

business process operations using add-on modules for SAS Enterprise Content Categorization. These add-on capabilities are scalable to huge quantities of information, providing the same processing speed for any amount of data while encoding that data in a highly compressed form. The add-on modules include text summarization, industry taxonomy rules for a quick start to building taxonomies, automated search and indexing, real-time content alerts, document duplication detection and a Web crawler. Add these modules to your SAS Enterprise Content Categorization solution as your needs evolve.

SAS® Text Miner

SAS Text Miner helps you discover information buried in unstructured text collections. The solution saves time and resources by automating the tasks of reading and comprehending text. The technology automatically parses text, identifies synonyms, creates frequency term counts, and can be mined to discover common topics and themes. As part of a fully integrated data mining environment, text topics and clusters can be included in descriptive or predictive analysis. Through interactive drill-down reporting and visualizations, SAS Text Miner empowers you to discover previously unknown patterns in document collections and apply those insights directly to your data mining models, back to your taxonomies or to corporate dashboards. By consolidating structured data sources with text-based information, you can observe trend details, spot new topics, send

alerts about potential problems and flag new business indicators more efficiently and with less risk. SAS Text Miner builds on the core data mining solution of SAS® Enterprise Miner™.

SAS® Sentiment Analysis

SAS Sentiment Analysis helps your organization to quickly understand customers' and constituents' opinions from across multiple sources of digital content – websites, communication centers, emails, forms, surveys, internal files and reports. It automatically locates and analyzes digital content in real time, then combines statistical learning with advanced linguistic rules to accurately reveal the core sentiments expressed in textual data. The software searches for and evaluates positive and negative phrases in text, including subtle emotional content. Using this detailed breakdown, you can create color-coded graphs to understand exactly what comments mean in relation to overall sentiment and changes in these sentiments over time. As a result, you can spot trends before they become viral, and identify your customers' priorities.

SAS® Ontology Management

Using advanced linguistic technologies to manage semantic terms, SAS Ontology Management associates text from different sources, file systems and topical areas. It enables you to create relationships among key information, including pre-existing and siloed taxonomies, so that subject-matter expertise about use and meaning can be systematically built into rules for

server processing and information search-and-retrieval systems. By consistently identifying relationships and using a standard approach to surfacing related content, queries will return relevant, comprehensive and accurate answers.

Why SAS®?

Speed and Improve Information Organization, Access, Findability, Knowledge Sharing

As globalization proliferates and the competitive landscape broadens, organizations are being forced to change from reactive to proactive. One of the key challenges with globalization is managing, collaborating and governing multilingual content. SAS Text Analytics enables organizations to fully harness text content assets. It also ensures reuse spanning disparate data repositories, regardless of who owns the content or where it was generated.

Purge Content Chaos

With SAS Text Analytics, you can abstract from a large set of documents across multiple enterprise repositories,

and focus on the content to recognize what has been duplicated or slightly modified and republished. This enables effective version control and better use of system and storage capacities through the entire content life cycle – from creation or acquisition through retention and disposition.

Discover Trends and Patterns

By applying natural language processing, advanced linguistic rules and statistical techniques to text, SAS Text Analytics can automatically transform textual data into usable information, finding explicit relationships or associations.

Cut Through the Noise of the Web and Other Social Media Outlets

With the ever-increasing volume and speed at which new content is published in today's information-centric world, it's critical to be aware of sentiment trends in real time. SAS Text Analytics helps you distinguish between positive and negative sentiments in multiple, sometimes conflicting product and service reviews by automatically extracting the true sentiments that are being expressed.

Count on an Integrated Business Analytics Framework

SAS provides you with an integrated business analytics framework that covers the entire intelligence creation process – from data integration and management, to advanced analyses and the efficient distribution of insightful information and reports through specialized business and industry solutions.

About SAS

SAS is the leader in **business analytics** software and services, and the largest independent vendor in the business intelligence market. Through innovative solutions, SAS helps customers at more than 55,000 sites improve performance and deliver value by making better decisions faster. Since 1976 SAS has been giving customers around the world THE POWER TO KNOW®.

www.sas.com