SPSS Inc.’s Clementine data mining workbench enables organizations to achieve measurable results by basing decisions on patterns and associations found in their data. But did you know that up to 80 percent of your organization’s data is contained in textual form?

The customer e-mails, call center notes, open-ended survey responses, Web forms, and other text sources that your organization captures—including content from RSS feeds, such as blogs and news feeds—contain up to four times as much valuable data as your organization’s structured data stores. This means that data mining projects focusing only on structured data may use less than 20 percent of the information available.

Text Mining for Clementine enables you to combine this valuable unstructured data with traditional structured data to significantly increase your understanding of customers, the public, and other groups. This product transforms Clementine into a fully integrated data and text mining workbench. You can perform both text mining and data mining within the interactive, visualization-based Clementine environment.

Powered by SPSS Inc.’s natural language processing (NLP) technology, Text Mining for Clementine pulls key concepts from many types of unstructured data and groups them into categories. Extracted concepts and categories are then combined with structured data and applied to predictive models to provide valuable insights into actions, behaviors, patterns, and associations.

In addition, Text Mining for Clementine uses text link analysis (TLA) to help you identify and extract sentiments, such as preferences and opinions, from text in multiple languages. By capturing the attitudes and thoughts of individuals across more channels and markets, you can gain a more complete view of your customers and constituents and build more accurate predictive models.
Add value throughout your organization

Text mining can be used in nearly any business or research situation that involves unstructured data. Here are just some examples of Text Mining for Clementine applications:

- **Product development and refinement.** Identify trends in complaints or requests by analyzing call center logs, customer e-mails, open-ended survey responses, and RSS feeds, including blogs. Use this information to improve existing products and services and develop successful new offerings.

- **Marketing campaigns.** Improve campaign effectiveness and revenue. For example, analyze inbound customer calls in real time to provide better product and service recommendations.

- **Churn prevention.** Discover why some customers leave—and take steps to prevent defection—by analyzing customer communication records for recurring problems or complaints that precede churn.

- **Cross-selling.** Improve sales by using information about customer preferences to better target products to existing customers.

- **Drug discovery.** Find relationships in chemical and biomedical databases.

- **Competitive intelligence.** Survey competitor and industry Web sites, RSS feeds, including blogs and news feeds, and other publicly available textual information to maintain a current view of your competitors.

- **Security.** Discover potentially suspect behavior by analyzing Web site content, chat rooms, e-mails, blogs, and other sources of information, and identify patterns and associations in the data.

Improve predictive quality and accuracy

The value of analyzing a combination of structured and unstructured data is both measurable and significant. A predictive model that is based on 100 percent of the available information is much more likely to provide accurate results than one based on only 20 percent or less of the data.

A mobile telecommunications provider, for example, used concepts extracted from its call center notes to improve the performance of existing churn models by 10 to 50 percent. A financial services organization—concerned about potentially non-compliant stock trades—created a model that tied information in internal e-mails to transactional data to predict which traders were most likely to break regulatory rules. By using Clementine and Text Mining for Clementine together, you integrate text mining directly into the analytical process, and ensure a measurable improvement in performance and results.

Unlock the power of text—no linguists required

Unlike other text mining products, you do not need a linguistic background or special training to use Text Mining for Clementine. And because Text Mining for Clementine uses an interactive interface within Clementine, text mining is straightforward and efficient. For example, interactive graphs enable you to explore and display text data and patterns for instant analysis.

You can easily customize concept dictionaries for a particular domain area by using the Resource Editor, an integrated resource for managing the text extraction process. This enables you to find relevant concepts and associations faster.
Create customized templates and libraries for specific business applications directly from the main Clementine toolbar. Text Mining for Clementine’s linguistic resources support a range of industries and applications, including sentiment analysis, CRM, security and intelligence, competitive intelligence, life sciences (genomics and mesh), and IT.

For example, a company that wants to analyze call center notes can use the Resource Editor to adjust Text Mining for Clementine’s dictionaries to reflect acronyms, abbreviations, and slang typically found in call transcripts. Or a pharmaceutical company can use the Resource Editor to set Text Mining for Clementine’s included genomics dictionary as its default.

**Deploy into operational systems**

In order to make the best use of your textual data, you need to be able to use it throughout your organization. Text insights deployed through Clementine predictive models to operational databases provide value to areas throughout your organization. And you can deploy any part of the text or data mining process by using Clementine Solution Publisher Runtime™.

To help manage your analytical assets and automate analytical processes, use Text Mining for Clementine with SPSS Predictive Enterprise Services™.

You can also use insights derived from text data to achieve more accurate results with SPSS Inc.’s predictive solutions. Improve real-time and batch scoring using PredictiveMarketing™, provide real-time recommendations to inbound callers through PredictiveCallCenter™, or improve claim processing with PredictiveClaims™.

If your organization collects customer insight through an enterprise feedback management (EFM) solution, Text Mining for Clementine can help you understand the opinions, attitudes, and preferences of your customers, employees, citizens, business partners, and others.

Text Mining for Clementine’s deployment capabilities make textual insights available to business users in critical areas, so your entire organization benefits from a comprehensive, 360-degree view of customers or, for government agencies, the citizens they serve.
**Text Mining for Clementine features**

**Methodology**
Clementine and Text Mining for Clementine support the CRoss-Industry Standard Process for Data Mining (CRISP-DM), which enables analysts to focus on solving business problems, rather than on programming. Text Mining for Clementine enables you to merge unstructured data with structured data during the CRISP-DM process. In addition, Text Mining for Clementine uses a proven combination of natural language processing (NLP) techniques and predictive analytics to efficiently extract meaningful information from unstructured data.

**Linguistic extraction**
- Extract text data from files or a database
- Work with multilingual text. Select from seven native language extractor options: Dutch, English, French, German, Italian, Portuguese, and Spanish
- Translate up to 14 languages using Language Weaver add-on options
- Manage errors in punctuation and spelling
- Extract domain-specific concepts such as unit terms, expressions, abbreviations, acronyms, and more
- Calculate synonyms using sophisticated linguistic algorithms and embedded or user-specified linguistic resources
- Name concepts by person, organization, term, product, location, and other user-defined types
- Extract non-linguistic entities such as address, currency, time, phone number, and social security number (SSN)
  - Templates for non-linguistic entities are available for all seven languages
- Included opinion, competitive intelligence, security intelligence, and genomics dictionaries enable relationship extraction; and the opinion dictionary also enables sentiment extraction (such as likes and dislikes).
- Cache extraction results during an interactive workbench session for reuse later

**Web feed node**
- Easily retrieve and analyze text from RSS feeds, such as blogs and news feeds, and HTML pages

**Text mining modeling node**
- Create clusters based on term co-occurrence using concept clustering algorithms, which provide an at-a-glance view of main topics and the way in which they are related
- Intelligently group text documents and records based on content, using text classification algorithms
  - Aggregates concepts from a wide variety of unstructured text data and groups them into a small number of categories
  - Reuses categories, enabling the scoring of any new text documents and records based on the text they contain
  - Accelerates and improves data management
  - Includes term inclusion and derivation lexical algorithms
- Enable advanced concept selection and deselection for use in subsequent Clementine predictive modeling applications
- List extracted concepts by type, frequency, document coverage, and other user-defined classifications
- Highlight synonyms used for each selected concept
- Convert selected concepts to structured form for use in Clementine predictive modeling algorithms
- Access all text mining capabilities through the text mining node palette in Clementine

**Text link analysis**
- Identify and extract sentiments (for example, likes and dislikes) from text in Dutch, English, French, German, Italian, and Spanish
- Identify links and associations between, for example, people and events, or diseases and genes

* Features subject to change based on final product release.  
Symbol indicates a new feature.

**Resource Editor**
- Support specific industries and applications using verticalized templates and libraries for sentiment analysis, CRM, security and intelligence, competitive intelligence, life sciences, and IT
- Create and edit custom libraries directly in the Text Mining for Clementine interface
- Define and edit:
  - Domain-specific terms
  - Non-linguistic entities
  - Synonyms
  - Concept libraries
- Edit the CRM, opinion, competitive intelligence, security intelligence, and genomics dictionaries

**Deployment**
- Deploy the entire data mining process automatically with Clementine Solution Publisher Runtime
- Send text mining extractions directly to Clementine for export into SPSS’ predictive solutions
- To help manage your analytical assets and analytical processes, use Text Mining for Clementine with SPSS Predictive Enterprise Services

**Open integration**
- Easily integrates external linguistic resources
- Accesses any text file or relational database via Clementine’s high-performance ODBC drivers
- Supports the industry-standard Predictive Modeling Markup Language (PMML)
- Supports standard text document formats, including plain text, PDF, HTML, Microsoft® Office, and XML
- Works with multilingual text natively (Dutch, English, French, German, Italian, Portuguese, and Spanish) and uses translations via Language Weaver options

To learn more, please visit [www.spss.com](http://www.spss.com). For SPSS office locations and telephone numbers, go to [www.spss.com/worldwide](http://www.spss.com/worldwide).

SPSS is a registered trademark and the other SPSS products named are trademarks of SPSS Inc. All other names are trademarks of their respective owners. © 2008 SPSS Inc. All rights reserved.  
TMC12SPC-0108