

Description:



The **SMART Text Miner** can identify terminology in context. A report shows multiple terms in a *key-word-in-context* format. The *easy-to-read* reports let users quickly identify and create the custom technical dictionaries.

The **SMART Text Miner** can process one file or multiple files in one pass. The user can set the number of occurrences for each term or word. A *Mining-at-a-Glance* feature shows one example for each word found in the text.

The left-context and right-context format quickly identifies nomenclature, jargon, measurements, poor style and common typographical errors.

The **SMART Text Miner** automatically ignores numbers, punctuation, graphics, XML and HTML tags. Multiple reports can be joined into one master report that show which document contains the term. The **SMART Text Miner** can write a report to a Word file or create an Adobe .pdf file. The text miner can process a *page in less than a second*.

A harvest feature links new terms from the text mining reports to the **SMART Lexicon Manager**. These reports help create the **MAXit Checker** dictionaries for a product or service.

First Release: April 2001 V1.1.0
Latest Release: Jan. 2020 V4.2.0
SMART Order: **SMART_MINER**

Operating Systems:

- Windows 7 (32-bit)
- Windows 10 (32-bit and 64-bit)

Hardware Requirements:

The minimum requirements: An Intel I7, 16 Gb memory and 1 Gb of available disk space.

One copy of the **SMART Text Miner** can be installed on a common server for a work group.

The **Text Miner** can process other languages. See a sample in Spanish to the right.

Text Miner STOP LIST:

The installation includes a *STOP LIST* that contains the high-frequency English words to ignore.

About the SMART Text Miner Software:

The software is developed by professional programmers in the USA. The software is written in Microsoft Visual Studio VC17 C/C++.

Highlights of the SMART Text Miner:

- **SMART Text Miner** processes either single files or groups of files in one pass.
- Multiple text file formats can be mixed in a group.
- Multiple reports can be merged and duplicate entries removed automatically.
- A 22-character **left context** and a 22-character **right context** format makes term identification faster.
- Automated term harvest feature for word lists.
- Easy-to-use interactive user interface.
- A knowledge of linguistics is not required.
- Software reads .txt, .xml, .sgml, .htm and ASCII text.
- A report shows the word or term frequency count and location in a document.
- A user can set the number of occurrences to record.
- The **Mining-at-a-Glance** feature reduces the report to show the **key words** in context.
- The installation includes a high-frequency STOP LIST and templates to customize stop lists.
- Links to **SMART Lexicon Manager** for dictionaries.
- The **SMART Text miner** identifies terminology, part names, materials, signal names, software, geographical references, people names, field names, abbreviations, acronyms and unique terminology.
- Output reports in .html, .xml and Microsoft Word.
- An interactive viewer feature allows the user to see the reports without the need to print the reports.
- Software can identify terms found in customer correspondence, maintenance and regulatory reports.
- Software can be programmed to read selected fields from large databases.
- A *term harvest* feature lets the user browse a report and move valid terms to the Lexicon Manager.
- Fast processing, less than a second per page.

Text Miner Report Format:

nes para nuestros	clientes	que son:
establecidos por la	comisión	Nacional Bancari
SAMPLE Spanish	complejas	y nuestros logros
con una trayectoria	comprobada	de diseño, imple
ro, sino que además	cumplirá	con los requisitos
ones validación de	datos	opcional, adminis