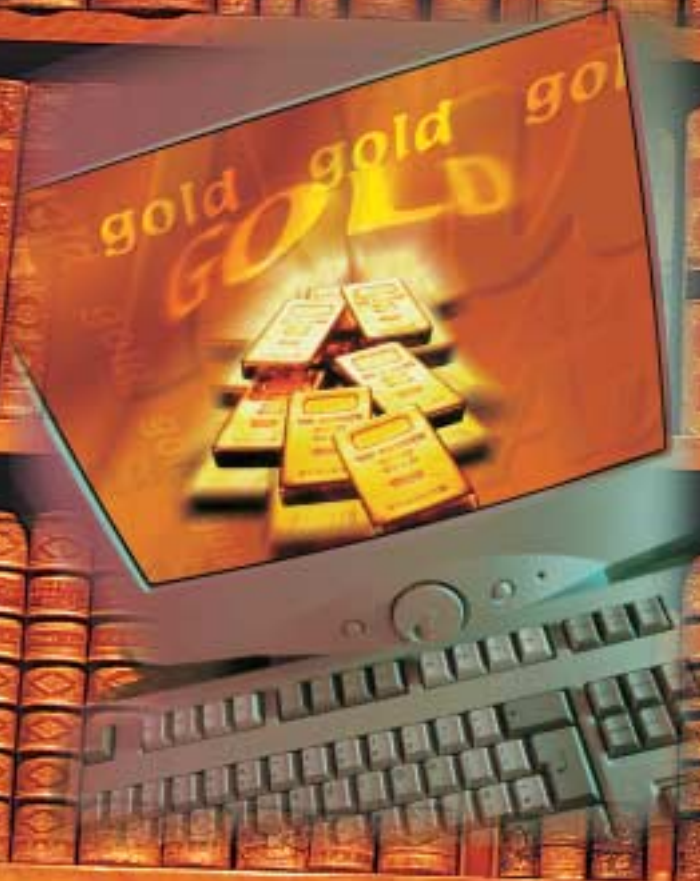


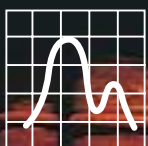
STATISTICA™

Text Miner *Software für den Erfolg!*

**Beherrschen Sie die Informationsflut
und erkennen Sie das Wesentliche in Texten!**



- ✓ **Umwandlung von Texten in auswertbare Informationen**
- ✓ **Zugriff auf Texte in diversen Formaten und über Websuche**
- ✓ **Auswertung von Texten in verschiedenen Sprachen**
- ✓ **Klassifizieren von Dokumenten, Kunden und Ereignissen**
- ✓ **Vollständige Integration in *STATISTICA Data Miner***
- ✓ **Automatisierung und Integration von Text Mining Lösungen**
- ✓ **Steuerung über das Internet**



StatSoft®

www.statsoft.de

Der *STATISTICA Text Miner* ist eine optionale Erweiterung des *STATISTICA Data Miner*. Er wandelt unstrukturierte Textdaten in eine verständliche und verwertbare Form, um Prozesse der Entscheidungsfindung zu unterstützen. In der Realität kommen Daten in vielfältiger Form vor und lassen sich dabei nicht immer direkt auswerten. Der *STATISTICA Text Miner* deckt dort verborgene Informationen auf.

Der *STATISTICA Text Miner* ist nahtlos in *STATISTICA* und den *STATISTICA Data Miner* integriert und bietet wie andere Produkte von StatSoft eine umfangreiche und mächtige Palette an Werkzeugen. Sie wurden mit Blick auf höchste Effizienz und Skalierbarkeit implementiert und setzen modernste Technologien ein, so dass hoch entwickelte Serversysteme mit multiplen Prozessoren optimal genutzt werden können.

Wie alle Komponenten des *STATISTICA Data Miner* wurde auch der *STATISTICA Text Miner* als universelles Werkzeug mit offener Systemarchitektur konzipiert. Die analytischen Werkzeuge des *STATISTICA Text Miner* lassen sich auf Textdokumente oder Webseiten anwenden. Aber auch aus vorverarbeiteten Bitmaps und Sounddateien heraus können unstrukturierte Informationen kategorisiert, gruppiert oder auf andere Weise in Analysen einbezogen werden.

Anwendungsbeispiele für den *STATISTICA Text Miner*

Inhalte von Webseiten analysieren.

Sie können Ihre firmeneigenen Internetseiten automatisch verarbeiten und zusammenfassen, um zum Beispiel Interessen und Trends in Besucherforen erkennen zu können.

Prädiktives Data-Mining mit unstrukturierten Texten.

Sie können Antworten auf offene Interviewfragen wie z. B. von Patienten beschriebene medizinische Symptome in Data-Mining-Projekte einbeziehen. Patienten und Symptome lassen sich so klassifizieren und diese Informationen zu prognostischen Zwecken verwenden.

Große Mengen an Dokumenten analysieren.

Sie können große Mengen an Dokumenten wie E-Mails von Kunden klassifizieren, um zum Beispiel Reklamationen anhand von Schlüsselworten zu identifizieren und zur effizienten Bearbeitung weiterzuleiten.

Integration in *STATISTICA*, *STATISTICA Data Miner* und *WebSTATISTICA*

Der Text Miner ist voll in die *STATISTICA*-Produktlinie integriert; es handelt sich nicht um ein allein stehendes Produkt, das von einem anderen Anbieter entwickelt und „irgendwie“ mit *STATISTICA* verbunden werden musste. Das macht diese Text-Mining-Lösung einzigartig: Durch die vollständige Integration stellt die Text-Mining-Funktionalität nur ein weiteres Modul dar, das in die Umgebung des *STATISTICA Data Miner*, von *WebSTATISTICA* oder von benutzerangepassten *STATISTICA*-Anwendungen integrierbar ist: Anwender können beispielsweise automatisch und routinemäßig auf Dateien, die in einem Data-Warehouse gespeichert werden, zugreifen, um bestimmte Analysen und numerische Zusammenfassungen der im Data-Warehouse verfügbaren Textinformationen zu aktualisieren. Diese Aktualisierungen können auch über *WebSTATISTICA* vorgenommen werden, so dass über das Internet weltweit auf die Ergebnisse dieser Analysen zugegriffen werden kann. All dies lässt sich mit der durchgängigen und intuitiven Benutzeroberfläche von *STATISTICA* realisieren.

Kernfunktionalitäten des *STATISTICA Text Miner*

Das Programm kann auf Textdokumente in verschiedenen Formaten zugreifen, einschließlich .txt (Text), .pdf (Adobe), .ps (PostScript), .html, .xml (Web-Formate) sowie auf die meisten Microsoft Office-Formate (z. B. .doc, .rtf).

Wie eine Suchmaschine unterstützt das Programm die Websuche, in dem es ausgehend von einer Startseite (URL) die verlinkten Dokumente aus dem Web extrahiert. Alle Dokumente, die mit dieser Seite verknüpft sind, sowie alle Dokumente, die mit diesen Unterdokumenten verknüpft sind, können bis zu einer benutzerdefinierten Tiefe einbezogen werden.

Der *STATISTICA Text Miner* enthält Ausschlusslisten und Wortstamm-Algorithmen für Deutsch, Dänisch, Holländisch, Englisch, Französisch, Italienisch, Portugiesisch, Spanisch, Schwedisch und weitere Sprachen. Die Ausschlusslisten können vom Anwender bearbeitet und erweitert werden.

Es gibt verschiedene Optionen, um die aus dem Text extrahierten Informationen in das *STATISTICA*-Datenblatt oder in externe Datenbanken zu übertragen.

Auf die numerischen Zusammenfassungen, die die Ausgangs-Texte repräsentieren, lassen sich grundsätzlich alle statistischen Analysemethoden anwenden.

Cluster-Techniken wie beispielsweise EM oder k-Means können angewandt werden, um Gruppen ähnlicher Dokumente zu identifizieren.

