

## Read Your Customers' Minds

Accelerate your business through text analytics



## Introduction to Text Analytics

### Unlock key customer indicators using text analytics

Analysis of textual content is one of the most critical business steps not typically implemented in today's data warehouse and business intelligence appliances. This is also true of integration of text analysis in business processes. As unstructured content becomes the core of our communication channels, text analysis solutions need to be adaptable so you can stay competitive. If you overlook text analytics, you could be missing a major opportunity to better meet your customer's demands.

RapidMiner offers innovative text analytics solutions that support all of your company's data needs where textual content is available, needs to be processed or can be analyzed. See how text analytics solutions can help reveal what your customers are thinking.

**Dr. Ingo Mierswa**, CEO and Founder, RapidMiner

### Textual data – a challenge for your company

Huge amounts of textual data are probably lying dormant in your company databases or are hidden in the vastness of the Internet. This is increasingly prevalent as more and more communication is shifting towards digital communication channels like e-mail and social networks. Most often, this textual data encompasses unfiltered statements made by a very important peer group – your customers. These statements, may they be complaints or even praise of your products, have one important thing in common: properly read, they can give you direct insight into your customers' thinking.

Despite this huge potential for gaining customer insight, textual data is still overlooked, rarely analyzed and seldom used in decision-making processes. This is due to the fact that the manual reading of texts and the extraction of insight is a tedious and time-consuming task. This task might be costly at best, but might even be impossible due to dramatically increasing data volumes. To overcome this issue, some text analytics software processes and analyzes textual data automatically, and can provide valuable insight by transferring the raw data into structured and manageable knowledge.

### Automate understanding and align analytics

Relevant textual data can emerge from various sources in- and outside your company and cover a wide variety of topics. Any opinion expressed or comment related to your product is inherently relevant since it allows you to gain insight into your customers' thinking. Text analytics solutions allow you to derive knowledge comparable to traditional market research, but in real-time as soon as the opinion is extracted. Moreover, those insights are available at a fraction of the costs of traditional market research. Consider solutions that offer automatic topic detection and sentiment analysis to extract valuable and structured information from text. Integration of these modules within your infrastructure means accessibility to the

information hidden within your data universe. Look for seamless integration within your data warehouses, co-existence with traditional KPIs and derivation of business intelligence, including customer insights.

### **Anticipate action through business process integration**

Posts in blogs, forums or social networks may be meant for a specific small community of people. The content however may get virally amplified and can flood the Internet in a matter of moments. Customers who send a complaint one day may be ready to quit your services the next day. Problems with products described in support tickets could be reported on various news outlets within hours.

Text analytics solutions can scan relevant statements automatically, monitoring complaints and support ticketing systems. Protect your company by providing added insurance against marketing catastrophes. The combination of text analysis with predictive modeling and forecasting lets you anticipate issues and act before it is too late and before you are forced in reaction mode, which comes with a higher price tag.

### **Route communication and save overhead**

The more successful your business is, and the more customers you have, the more you interact with your customer community. But this success comes at a cost because of the increased level of communication necessary to be efficient and react in a timely manner. Wouldn't you like to reduce the amount of emails that need to be manually read to 20 percent? Or reduce the number of times an email needs to be reviewed until it reaches the right expert? Look for automatic text categorization to support email routing and send support tickets to the appropriate recipients. Real-time processing reduces latencies to a minimum. Not only can you save time and money but the customer experience will be improved.

### **Evaluate content, target communications, control impact**

Without a doubt, the volume of content exchanged between you and your customers will continue to increase in the coming years, yet it is unlikely the quality of this transmitted content will improve. With text analytics the gap between high volume and poor quality can be closed. You do not need to waste time manually reading and evaluating content. Going forward, you will only read what is relevant to your business. Are you thinking of your customers and leads as well? Maybe you have subscribers that might be bothered by potentially irrelevant information sent by your organization? Text Analytics allows you to learn what is relevant to each of your subscribers and to send only the content that interests them. This means true targeted communication, with you in control of the impact.

### **Recommendations**

- If text analytics serves a strategic purpose in your organization, create a centralized program to enable an efficient and coordinated approach to implementation.
- If text analytics are only for specific tactical purposes look for niche, "best-of-breed" vendors to support different use cases rather than trying to find products that can solve all problems or attempting to force unsuitable products do things they aren't meant to do.
- Acquire the necessary skills for the effective deployment of text analytics and assign people trained in the use of these technologies.
- Combine text analytics and Business intelligence tools for even larger benefits such as hybrid analytics.

## GfK uses RapidMiner solutions for the analysis of Internet and social media text

*Enterprise: GfK Marktforschung, digital research department*  
*Branch of industry: Market research*

### Situation

The GfK group is one of the leading market research enterprises worldwide, operates in more than 100 countries and with more than 11,000 employees. GfK provides services for all major consumer goods, pharmaceutical, media and service markets. These services are divided into two sectors: Consumer Choices and Consumer Experiences. Consumer Choices supplies data reflecting consumer decisions and activities. Consumer Experiences is concerned with consumer behavior and attitudes and how people perceive and experience the world.

Since the Internet is playing an ever greater role in the business life of today, the collection and analysis of complex online data are becoming increasingly important. Online market research in terms of surveys carried out over the internet has existed since the 1990s. Now the Internet itself is being surveyed – and the subject is the traces of communication left behind on it by users, e.g. in the form of comments within a forum for a certain brand, enterprise etc. In order to benefit from this unstructured information, the data needs to be categorized, using, for example, sentiment analysis, which collects sentiments and opinions expressed on the internet.

GfK acknowledges the increasing significance of information on the Internet and entrusted the digital research department with the evaluation of this data. This requires text data be collected from the Web or from surveys and evaluated. Taking a wide variety of data sources into account, pages on the Internet need to be searched (crawled) and content extracted and analyzed. On the basis of this data analysis, reliable statements should then be possible e.g. concerning the attitude of Internet users towards certain products. An adequate analysis solution was sought in order to cope with the increasing amount of texts from online sources.

### Challenge

The digital research department knew exactly what it was looking for -- an analysis solution for evaluating Web content. The solution needed to offer machine-based learning, e.g. in the form of automatic text categorization. A generic solution with high adaptability and reusability was sought, one which could

**"We looked at a number of other products on the market, but none were as able as RapidMiner to meet the requirements for flexible and sometimes short-notice use with reliable support."**

Thomas Eggebrecht  
senior IT specialist  
at GfK Marktforschung

**"Thanks to the simple handling, the low software and hardware requirements, the ease of integration and last but not least the possibility of worldwide collaboration, we are optimally equipped to process more or less any request quickly and competently following web content analyses."**

Thomas Eggebrecht  
senior IT specialist  
at GfK Marktforschung

perform an indicative evaluation of the information found in user generated content (UGC). Moreover, the requested data had to be accurately extracted from unstructured information found on websites and given the necessary meta data (e.g. publication date). Data analysis from social media sources plays an ever increasing role

The solution needed an intuitive interface, since different groups of employees with varying levels of knowledge (IT specialists, research consultants or analysts) had to be able to use it. Speedy computation of large quantities of data had to be possible on "small" laptops. The uncomplicated exchange of analysis processes between employees had to be ensured. The processing of typical data formats from market research (SPSS, MS-Office, ASCII, txt) and the integration with SQL databases was also essential.

## **Solution**

After a thorough market evaluation, GfK picked RapidMiner Studio. As time went on, even more enormous volumes of data needed to be evaluated, so GfK added RapidMiner Server, a high-performance analysis server. This provides greater ease of integration and interactive visualizations as well as higher performance. The possibility of replication and reusability was a decisive factor for choosing RapidMiner. Other products that had previously been considered were ultimately rejected for cost reasons or due to restrictions in functionality, a lack of support or insufficient system openness.

Used together with RapidMiner Server, RapidMiner Studio also functions as a kind of user interface: users access RapidMiner Server via RapidMiner Studio, and analysis processes then run 24 hours a day in the background so that performance is not compromised and a high number of simultaneous users can be served. RapidMiner Server enables the use of much more powerful hardware and more working memory thanks to a client-server architecture, makes collaboration possible and improves collective working. Since the files are stored on the server in the repository, other users can work on the data and you are spared the hassle of transferring them from one computer to the other using a USB stick (as was still the case with the local version of RapidMiner Studio). Otherwise the high volume of data could hardly be coped with.

RapidMiner Studio scored points with its simple user interface, reasonable price and comprehensive support. The fact it is an open source solution was an important factor from the beginning. For this means temporary users such as external suppliers or temps can use the system without a license and at short notice.

"The open source thought was important when choosing a high-performance and equally cost-efficient data analysis solution," points out Thomas Eggebrecht, senior IT specialist (head of programming) at GfK Marktforschung. "We looked at a number of other products on the market, but none were as able as RapidMiner to meet the requirements for flexible and sometimes short-notice use with reliable support."

Another advantage of RapidMiner Studio is the high degree of flexibility: The solution runs on all systems thanks to Java, and the exchange of analysis processes between employees is ensured by way of XML files. There is a simple update mechanism and processes can be executed by script both under Linux and under Windows. With its openness, the solution can also be extended at any time to include its own plugins or operators at the open source Java API. What is more, all usual file formats in market research are supported. RapidMiner Server also offers the unsupervised and job-controlled running of long-lasting processes on a server. Since a virtually unlimited number of users can access the data via the analysis server, work between several persons on a project is made easier and data is exchanged via a remote repository. Using the RapidMiner solution does not make any special demands on the hardware or software either, since a normal Linux server with Java RE is used and so no complicated installation routines or root rights are necessary.

## Simple installation

The introduction of RapidMiner Studio did not involve a trial phase; decision-makers were able to try out the solution for themselves at a training event at RapidMiner. It was possible to install RapidMiner Studio in just a few minutes – and no special know-how was necessary. RapidMiner Server required an IT specialist to come in for one day for the server setup and the installation, created users and integrated the remote repository into RapidMiner Studio.

Want to learn more about how RapidMiner's solutions for Text Analytics can accelerate your business?

Send us an email at [contact@rapidminer.com](mailto:contact@rapidminer.com)  
or call us at +1 617-401-7708  
or +49 231-425-786-90.

## International collaboration

At GfK, data is collected from web texts with RapidMiner Studio. Internet sites are searched using crawling processes and content is extracted (advertising weeded out, purged of HTML structure), which is then stored in the data warehouse. A sentiment analysis is then carried out on the data. German websites are searched here as well as international ones. While RapidMiner is used at GfK in Germany, the projects are international in scope.

GfK saves a huge amount of time and effort with a generic content extraction model, which extracts relevant data from almost any Internet sources. This notably makes it possible to reuse processes or process elements and save them as a template or library, so that they do not have to be rewritten for each crawling action. In this way processes run automatically and manual interaction by the user is no longer required.

## Outlook

"GfK can offer its customers high-quality, controllable and comprehensible methods with the RapidMiner solutions," explains Eggebrecht. "Thanks to the simple handling, the low software and hardware requirements, the ease of integration and last but not least the possibility of worldwide collaboration, we are optimally equipped to process more or less any request quickly and competently following web content analyses."

Eggebrecht expects a high demand for text mining and data preparation for the future, including from other sectors such as retail and technology. "After the positive experience introducing the RapidMiner solution, I could definitely imagine it being used worldwide," says Eggebrecht. "GfK is a global enterprise and our projects are international, so the region-wide use of uniform software makes good sense."

---