



# Text Analytics: The Hurwitz Victory Index Report

SAS



Fern Halper, Ph.D  
Partner and Principal Analyst

Marcia Kaufman  
COO and Principal Analyst

Daniel Kirsh  
Senior Analyst



**HURWITZ  
& ASSOCIATES**  
Insight to Action



**Table of Contents**

Introduction 3

    Market Trends  
    for Text Analytics 3

What is the Victory Index? 6

Victory Index Methodology 6

    Key Metrics 6

    Data Sources 8

    Vendor Selection 9

Results 9

    Market Perspective 10

        Victors . . . . . 10

        Leaders. . . . . 11

    Customer Perspective 12

        Victors . . . . . 12

        Special Mention-Value. . . . 13

        Special Mention-Validity . . 13

        Leaders. . . . . 14

        Challengers . . . . . 14

    Vendor Assessment: SAS 15

## Introduction

Unstructured data is everywhere – in emails, call center notes, verbatim survey responses, tweets, blogs, and online news - to name just a few sources. In fact, by most industry estimates, unstructured text accounts for eighty percent of the data available to companies. Text analytics is being embraced at a rapid rate by organizations that want to gain insight into this information and use it for a competitive advantage. Factors fueling this growth include a better understanding of the technology’s value, a maturing of the technology, the rise of social media as a source of potential insight, and the compute power to help analyze large amounts of data.

In fact, text analytics has become a key component of a highly competitive company’s analytics arsenal. Hurwitz & Associates defines text analytics as:

Text Analytics is the process of analyzing unstructured text, extracting relevant information, and transforming it into structured information that can be leveraged in various ways.

The analysis and extraction processes used in text analytics take advantage of techniques that originate in computational linguistics, statistics, and other computer science disciplines. The transformed information from text analytics can be combined with structured data (e.g., sales and demographic data) and analyzed using various business intelligence or predictive and automated discovery techniques. Or, the text can be extracted and transformed and then analyzed iteratively to determine relationships and trends, look for clusters, and other information. Organizations also use text analytics to help enhance search, for example, in faceted navigation.

Table 1, below highlights some of the popular uses for text analytics.

**Table 1. Use cases for text analytics**

Area	Use Case
Marketing	Voice of the customer, social media analysis, churn analysis, market research, survey analysis
Business	Competitive intelligence, document categorization, human resources (voice of the employee), records retention, risk analysis, website faceted navigation
Industry specific analytics	Fraud detection, e-discovery, warranty analysis, medical research

## Market Trends for Text Analytics

As end users grow to understand the value of text analytics, the market itself is evolving to deal with a variety of data issues and different deployment and

*Text analytics is being embraced at a rapid rate by organizations that want to gain insight into this information and use it for a competitive advantage.*

delivery models. Hurwitz & Associates sees the following trends in the text analytics market:

- **Trend 1: Social media analytics adoption drives text analytics.** Social media is definitely a driving force behind the adoption of text analytics. There are more than one hundred vendors that provide social media monitoring platforms with varying degrees of sophistication in terms of the analysis. A number of them have moved beyond search-based techniques and include text analytics to discover topics or analyze sentiment. Several of the vendors profiled in this report including Attensity, Clarabridge, IBM, and SAS, offer social media analytics utilizing text analytics. Many current adopters start analyzing unstructured text for social media and then move beyond this to explore other use cases.
- **Trend 2: Analytics moves beyond sentiment analysis in text projects.** Sentiment analysis is viewed by a large percentage of the market as a key feature for text analytics solutions, although most people realize that it has its limitations in terms of accuracy. Companies want to know what people think about their product/service compared to their competitors or how their customers feel about a certain topic. Sentiment analysis is evolving. Vendors like Clarabridge offer sophisticated sentiment on an eleven-point scale, rather than simply classifying a document or a phrase as positive, negative, or neutral. SAS can monitor sentiment and evaluate changes over time. For example, its software can determine whether sentiment during a certain time period was more negative, than say a previous time period. Vendors are also moving beyond sentiment as the analytic for a text analytics solution. For instance, IBM recently announced its “birth of a trend” analysis. Its software has the capability to extract topics, which grow or diminish over time. The end user can either define a set of topics for the system to monitor or the system can intelligently identify emerging topics from a set of data sources. Other vendors, such as Clarabridge offer an analytic called “emerging issue” which utilizes auto classification techniques.
- **Trend 3: The market begins to get the connection between text and Big Data.** As companies gather greater volumes of disparate kinds of data (i.e. structured and unstructured), they are looking for solutions that can scale to analyze this increased volume. For example, companies might use this capability to analyze real time news feeds for financial reasons. Or, a company might analyze comments about its brand in social media for real time ad placement. The buzz from text analytics vendors about “Big Data” solutions is growing louder too. For instance, companies like Attensity and SAS are providing grid solutions and massively parallel processing techniques to help companies analyze real time unstructured (and structured in the case of SAS) streams. Attensity, IBM, OpenText, and SAS have embraced Hadoop and MapReduce. It makes intuitive sense that text analytics vendors would embrace Big Data technologies since unstructured data is a big part of Big Data. On the other hand, vendors need to be cautious about the Big Data moniker in their marketing campaigns since they may confuse the market.
- **Trend 4: Marrying structured and unstructured data becomes more popular in analysis.** End users realize that they can gain significant insight from mining unstructured (i.e. text) data, and that this unstructured data, when used in conjunction with structured data, can provide a big lift to predictive models. Some companies, such as IBM, Megaputer and SAS have been providing this ability for a number of years. Companies such

*...unstructured data, when used in conjunction with structured data, can provide a big lift to predictive models.*



as Lexalytics are providing this capability in an OEM capacity to predictive analytics vendors like Angoss Software, which recently announced it now had the capability to add unstructured data to its decision tree analysis. Of course, marrying structured and unstructured data together can be useful in many types of analysis, not just predictive analysis.

- **Trend 5: Integrated information management now includes text.**

Increasingly, end users are becoming aware that integrating disparate data (e.g. structured data with unstructured data or unstructured data with other unstructured data) is becoming increasingly important. Integrating data, data quality, and governance are becoming critical concerns. OpenText recently announced its InfoFusion platform that federates information across unstructured sources for the purpose of accessing, aggregating, analyzing and managing data. Over the next several years, the company also plans to create an integrated data model for unstructured data. Other vendors, such as SAS offer ETL capabilities that can be used with structured and unstructured data.

- **Trend 6: Text analytics continues to enhance search.** When companies first start to think about text analytics, they often think search, although search and text analytics are different. Keyword search is about retrieving information based on what end users already know they are looking for. In contrast, text analytics is generally more about discovery. However, text analytics is being used to enhance search. For example, the OpenText Semantic Navigation uses OpenText Content Analytics technology to index web pages and social media content for navigation via semantic facets. IBM Content Analytics with Enterprise Search utilizes converged content analytics and enterprise search targeting line of business analytics and search use cases. The vision is broader. TEMIS, for example, is at the early stages of offering a linked network of content at the entity or concept level. IBM offers Watson, which can function as a question and answer system. Watson works by generating hypotheses, which are potential answers to a question. It is trained by feeding question and answer data into the system.

- **Trend 7: Content Management Systems (CMS) leverage text analytics to automate classification of documents.** When Hurwitz & Associates performed one of the first studies on text analytics back in 2007, we saw that companies wanted to integrate text analytics with their content management systems. More recently, vendors are now providing this feature, especially to help automate the classification of documents, in content management systems. OpenText, for example, can layer several levels of semantically extracted or inferred metadata that can leverage each other to create new annotations, in near real-time, without the need to re-index or re-compile the underlying data structure. TEMIS has a partnership with EMC Documentum. IBM Content Analytics with Enterprise Search works with FileNet.

- **Trend 8: Text analytics delivered on mobile.** Just as mobile is a hot technology in business intelligence; many text analytics vendors have also made strides to deliver text analytics on mobile devices. For example, the Lexalytics engine has been ported to Android, and can be used to provide native-text processing within Android devices. The front end of OpenText's Auto-Classification is a Web app, designed for mobile devices. Attensity Analyze includes a tablet application that is now available for smartphone devices. Through SAS Mobile BI, SAS Text Analytics output to mobile devices is supported. Additionally, the Text Summarization add-on to SAS Enterprise Content Categorization can be used to reduce document contents to only

*Just as mobile is a hot technology in business intelligence; many text analytics vendors have also made strides to deliver text analytics on mobile devices.*



the key sentences contained within the material for clarity in messaging devices.

- **Trend 9: The cloud becomes more popular for text analytics.** In addition to social media helping to drive the adoption of text analytics, so too is the Cloud. In fact, a number of the vendors profiled for this report offer various cloud options for text analytics (see vendor descriptions below for details). Often, companies feel that they don't have the in-house skills to aggregate and analyze unstructured information. They are turning to Software as a Service (SaaS) solutions for help. In fact, this is one of the fastest growing delivery models for text analytics. There are also newer models emerging for text analytics in the Cloud. For example, Pragmatech, an emerging company to watch, is offering an API semantic engine, which it views as a semantic enabler that can be plugged into many applications. It will be available in a cloud model.
- **Trend 10: Open source solutions for text analytics grow.** There are many open source vendors providing text analytics solutions. Open source solutions are becoming increasingly important to the text analytics market because they enable a large and diverse community to engage in innovation.
- **Trend 11: The notion of semantic enablers begins to gain momentum.** The notion of semantically enabling an application is also gaining steam. Some companies offer "semantic middleware" (e.g. OpenText) while others offer "semantic APIs." (e.g. Alchemy API and Pragmatech). The idea is for the end-user to be able to enable semantic analysis where it is needed.

These are only some of the trends that are making text analytics a vital part of the analytics market landscape.

## What is the Victory Index?

The Hurwitz & Associates Victory Index is a market research assessment tool, developed by Hurwitz & Associates that analyzes vendors across four dimensions: Vision, Viability, Validity and Value. What is unique about the Victory Index? Hurwitz & Associates takes a holistic view of the value and benefits of important technologies. We assess not just the technical capability of the technology but its ability to provide tangible value to the business. This Victory Index focuses on text analytics. Intensive research by our analysts on the business value of this technology supports our conclusion of great potential for continuing innovation and market growth in this arena.

## Victory Index Methodology

### Key Metrics

In order to provide a comprehensive understanding of the market and the vendors, Hurwitz & Associates applies a rigorous methodology that employs approximately 50 attributes across four dimensions, which are analyzed with a weighted algorithm. These dimensions are explained below and the specific metrics are presented in Table 2.

*Often, companies feel that they don't have the in-house skills to aggregate and analyze unstructured information. They are turning to Software as a Service (SaaS) solutions for help. In fact, this is one of the fastest growing delivery models for text analytics.*



- **Vision: The strength of the company’s strategy.** Attributes evaluated here include: Is the vision not only realistic but also compelling? It is not enough to simply have a good vision; a company must also have a well-designed road map that can support this vision. Vision attributes also include more tactical aspects of the company’s strategy such as a technology platform that can scale, well articulated messaging, and positioning. A key component of this dimension is clarity: it must be clear what business problem the company is solving for which customers.
- **Viability: The strength and vitality of the company in the market.** There is often not a direct correlation between the market strength of a company and its revenue or the number of years that it has been in business. This is especially true when a market is emerging. Sometimes an emerging company can become a major force within a few years if they solve a complicated customer problem. Hurwitz & Associates evaluates a complex set of financial and company-specific attributes that, when combined, help to determine the viability of a business. Attributes include financial ratios, customer adoption rates, intellectual property, strength of management team, and strength of partnerships. We are also concerned with the vitality of the company in a particular market. For example, a large company might be strong and vital in a certain product area, but may be lagging in the text analytics space because of limited investment, a poor market push, or stagnant products that have seen little change or innovation for years.
- **Validity: The strength of the product that the company delivers to customers.** There is a distinction between the positioning of a product and its ability to satisfy customer requirements. Therefore, the Victory Index analyzes how well a company executes on its promises. This part of the Index examines the features and functions delivered. It looks at the depth of functionality as well as the effectiveness of a product at evolving based on changing customer requirements. Other important attributes include ease of use, innovation, how well the product integrates with other technologies, and how well it incorporates important de facto and de jure standards.
- **Value: The advantage the technology provides to customers.** Even if a product is well designed, it must be able to help businesses achieve their goals. Goals range from gaining insight about customers in order to be more competitive, to using the technology to increase revenue. A key attribute that is measured in this dimension is how well the product supports companies in meeting their objectives.

*Even if a product is well designed, it must be able to help businesses achieve their goals. Goals range from gaining insight about customers in order to be more competitive, to using the technology to increase revenue.*

**Table 2: Victory Index attributes**

**Dimension Attributes**

Vision	<ul style="list-style-type: none"> <li>• Vision: clarity of vision, practicality of vision, a compelling vision, target market alignment, product direction and road map, leadership team vision</li> <li>• Messaging: strength of message, key differentiator value</li> <li>• Direct feedback from customers regarding vision</li> </ul>
Viability	<ul style="list-style-type: none"> <li>• Viability measures: revenue/employee, net profit, current assets/ current liability, revenue growth, customer growth</li> <li>• Vitality: strength of intellectual capital, strength of leadership team, customer adoption, customer stability, competitive strength</li> </ul>



	<p>in a particular market, market presence, channel strategy, innovation</p> <ul style="list-style-type: none"> <li>• Direct feedback from customers regarding viability/vitality</li> </ul>
Validity	<ul style="list-style-type: none"> <li>• Direct customer measures: breadth and depth of functionality, technical benefit, tools and technology, technical value, performance, scalability</li> <li>• Product functionality: support for ontologies and taxonomies, sentiment accuracy, integration with other systems and data sources, language support, visualization support</li> </ul>
Value	<ul style="list-style-type: none"> <li>• Direct customer measures: customer satisfaction, customer support, ease of use, product innovation, business benefit as determined by customer, time to value, value/price</li> <li>• Social media metrics around customer loyalty and brand value</li> </ul>

The Vision, and Viability indicators are further grouped under the heading of “Market Strength” since these metrics, analyzed together, give a good indication of the strength of the offering to the market. The Value and Validity measures are also analyzed together since they give a good indication of how customers view the offerings. These are grouped under the heading of “Customer View.”

### Data Sources

Data for scoring comes from multiple sources:

- **Vendor surveys:** An extensive survey was designed to obtain information on each vendor’s vision, strategy, products, financial, and company stability, as well as what the company believes to be its value. Hurwitz & Associates also conducted briefings with each vendor to further augment and understand this information. In several cases where the vendor chose not to complete a survey, information obtained from other data sources were compiled in order to understand these attributes and a written summary was provided to the vendor for fact check and review.
- **Vendor references:** Each vendor supplied 3-5 names of clients who provided input regarding the value and validity of products to Hurwitz & Associates. Hurwitz & Associates either sent a survey to each of these clients, conducted a phone conversation with them, or both. Given that some companies use more than one text analytics solution, these conversations were also used to understand where one vendor’s product provided value in comparison to other vendor solutions.
- **Online Victory Index Survey:** In addition to speaking with vendor references, Hurwitz & Associates conducted several online surveys to gather direct customer feedback regarding vendor products. We compiled results from more than one hundred and fifty companies regarding how they were using vendor products, what they were using them for, and their experience with the product. We also asked respondents to rate the vendor’s products across the Victory Index dimensions. Survey results provided an objective view of each vendor’s offering.
- **Social media:** A unique aspect of the Victory Index is that it uses social media data to help inform results. Social media was used, when appropriate, to determine brand media dominance, brand loyalty and strength, and brand

*Hurwitz & Associates compiled results from more than one hundred and fifty companies regarding how they were using vendor products, what they were using them for, and their experience with the product.*



value. The commercial product, NetBase ([www.netbase.com](http://www.netbase.com)), was used to perform this analysis. Netbase Social Media Insight & Analysis is used by marketing teams to understand online brand equity, analyze and compare customer passion, and generate insights that answer “Why” questions. The platform is used by hundreds of corporate customers, and was developed in partnership with five of the top 10 consumer-packaged goods companies, including Coca-Cola and Kraft.

- Hurwitz inquiries: Input from discussions with Hurwitz clients was used in the study.
- Other secondary sources included investment reports and public filings among others.

### Vendor Selection

Text analytics is a relatively new and emerging market and is therefore fragmented. Some text analytics vendors are particularly strong in one business area such as the voice of the customer. Some have traditionally focused on survey analysis, but are moving into other areas. Others are more general purpose. We felt it was important to present buyers with a range of potential solutions from both large and small players.

We also decided not to include social media analytics companies per se in the Victory Index for Text Analytics although many do provide text analytics capabilities. We felt that it would confuse the reader because although most of the vendors in the text analytics market can compete in the social media analytics market, text analytics vendors have far more offerings. We also did not include specific vertical text analytics solutions.

We believe that all of the selected vendors are strong contenders in the Victory Index.

The companies profiled in this study include Attensity, Basis Technology, Clarabridge Inc., IBM (Content Analytics with Enterprise Search and SPSS), Lexalytics, Megaputer Intelligence, OpenText, Provalis Research, SAS, and TEMIS.

In our online survey, we received customer feedback about other companies including Alchemy API, Anderson Analytics (Odin Text), and Verint. However, there was not enough information to formally include these companies in this report in terms of vendor profiles. Additionally, SAP and Oracle (which recently purchased Endeca), which both have text analytics capabilities, are also not included in this report.

### Results

Results are based on a thorough analysis of the data described above. Vendor scores result from use of a weighted algorithm across the four dimensions of the Victory Index. The results appear in two charts. Both are scatter plots, which illustrate where each vendor scored relative to the median scores of the group. Figure 1 (Market Strength) is a plot of Vision vs. Viability. Figure 2 (Customer

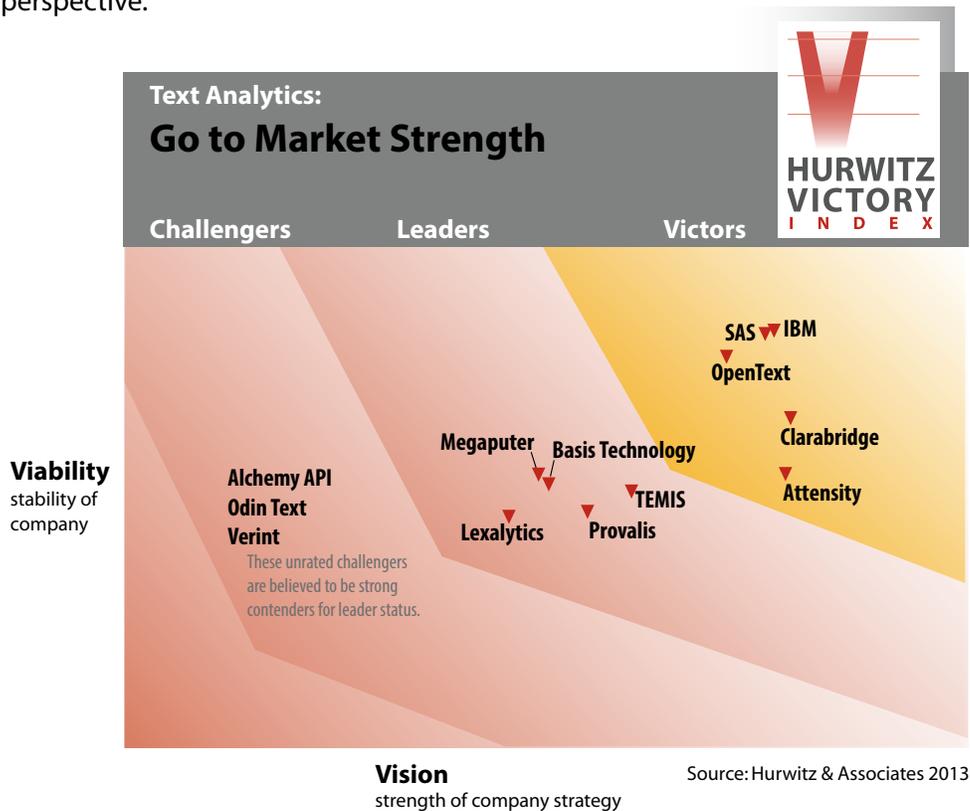
*Vendor scores result from use of a weighted algorithm across the four dimensions of the Victory Index.*



View) is a plot of Validity vs. Value. Complete vendor profiles appear in the next section.

### Market Perspective

Figure 1, below, illustrates the results of the Victory Index from the market perspective.



*The Victors demonstrate significant strength in brand and have a strong market presence in text analytics.*

### Victors

The Victors demonstrate significant strength in brand and have a strong market presence in text analytics. Their financials indicate stability, they have a compelling vision and strong strategy for their technology, and their customers resonate with this vision. Their customers believe these companies are innovative in the text analytics space and their solutions provide strong value. Each Victor has a combined Vision and Viability score that ranks at or above the median scores across all the attributes measured in these two categories. IBM, SAS, OpenText, Clarabridge and Attensity are all Victors. IBM and SAS are the top two market Victors.

- **IBM.** IBM has developed a strong strategy and vision for analytics and text analytics is an important component for its solutions as part of its Smarter Planet strategy. Smarter Planet holds a vision of the world as a more interconnected, instrumented, and intelligent place. IBM has a solid leadership team in place and a roadmap to execute on its strategy. IBM has a strong customer base that it can continue to grow as it provides a broad range of analytics capabilities including predictive analytics that utilize unstructured information. It also has a strong industry focused solutions group as well as an impressive research team. It will continue to integrate its

text analytics solutions and innovate. IBM solutions should be able to tackle a range of customer problems.

- **SAS.** SAS has put a great deal of thought into assembling a compelling suite of strong text analytics related products. Its purchase of Teragram several years ago to solidify its text analytics capabilities was a good move for the company. Additionally, SAS' approach to developing an analytics platform that can handle unstructured data makes sense. SAS is strong financially and its customer base continues to grow. It has a solid roadmap for the future and continues to innovate in a number of areas, such as Big Data and analytics, which will provide customer value.
- **OpenText.** A leader in content management, the OpenText strategy has evolved to address Enterprise Information Management, which involves managing, securing, and extracting value from unstructured data. Given this, the purchase of Nstein several years ago to help power this vision was a sound one. Customers also seem to resonate with the OpenText strategy and the idea of semantic enablers – i.e. that semantics can be utilized where needed. The company is in a growth mode and seems to be delivering solid results.
- **Clarabridge.** A historically pure-play text analytics vendor, Clarabridge is now firmly focused on customer experience management. It has been growing steadily and is getting very good traction with its SaaS offerings. Its customers resonate with its vision and believe the company does a good job innovating and giving them the software they need to succeed.
- **Attensity.** Another historically pure-play vendor, Attensity has a solid vision around leveraging multichannel voice of the customer conversations for business advantage. With strong technological expertise and growth from its acquisition of Biz360, it is a leader in the social CRM space. While focused on social CRM currently, the company also has the technology to succeed in other areas where text analytics is (and will continue to be) important to users.

## Leaders

Leaders demonstrate solid brand and financial stability and have a good vision for their text analytics solutions. However, companies in this category did not achieve Victor status for two main reasons. First, some Leaders are highly viable organizations, but their vision and strategy for text analytics is not as clearly articulated and well executed as the companies rated as Victors. Second, their growth or growth trajectory may not be as strong as the Victors.

- **TEMIS:** TEMIS, which has a solid customer base in Europe, has a very interesting vision around linkage networks. This is a network of semantic links between documents, which can power new forms of faceted navigation and provide context by linking related documents and knowledge. Its success with this vision will depend on how quickly and how fully it executes on this vision.
- **Basis Technology:** Basis Technology provides a very strong solution for customers looking for multi-lingual support. As part of its growth plan, it recently opened an office in Europe. Primarily known as an OEM player, the company is working on a strategy to bring more solutions to the text analytics space.

*Leaders demonstrate solid brand and financial stability and have a good vision for their text analytics solutions.*

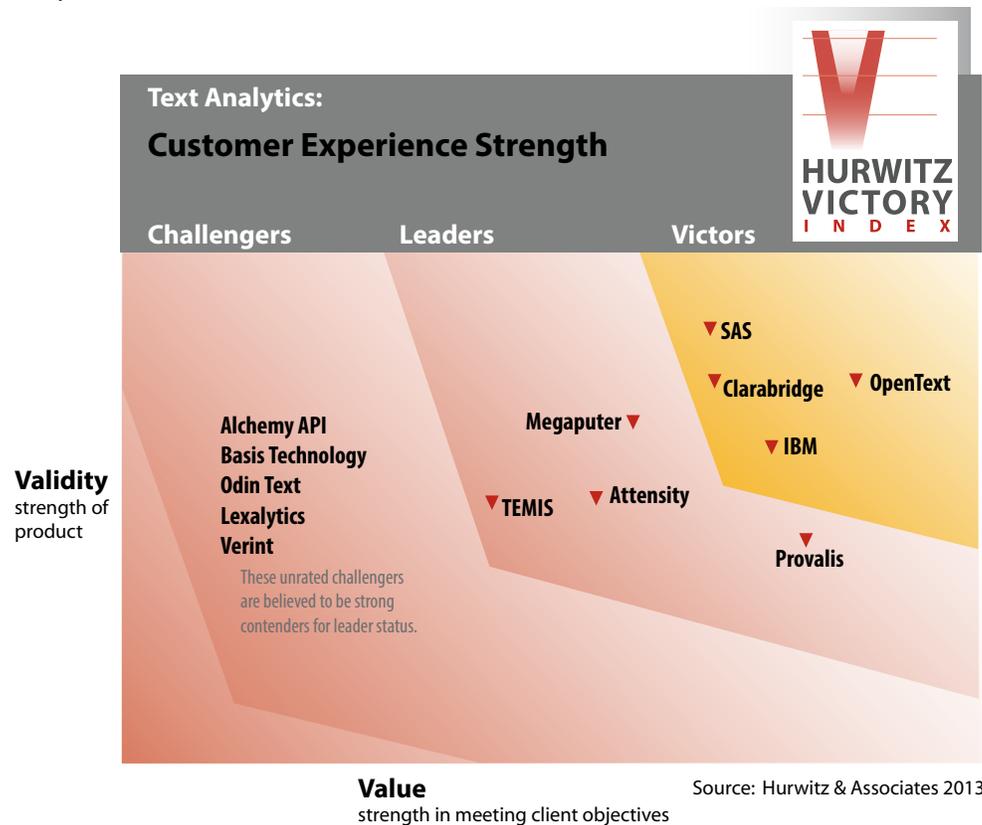


- **Provalis Research:** Provalis Research’s goal is to offer a range of affordable and flexible tools that can help those with both small and large text analytics projects. Its customers really like the value for the price (see below). Since it is a small company, it will need to ramp up to execute its vision to grow significantly.
- **Lexalytics:** After divesting itself from its 2008 Infonics merger, Lexalytic’s vision and strategy is squarely focused on the OEM market. Its goal is to be a leader in this market and it has won some impressive contracts. It will be interesting to see how it moves forward with its vision.
- **Megaputer Intelligence:** Megaputer Intelligence, with its service focus on analyzing both structured and unstructured data separately or together has a good strategy. Customers have a positive view of the company’s technology. Based on the impressive wins the company achieved in 2012, we expect Megaputer will attain Victor status as it continues to build market presence.

SAS, Opentext, IBM and Clarabridge are all Victors in this category.

### Customer Perspective

Figure 2, below illustrates the results of the Victory Index from the Customer Perspective.



### Victors

The Victors demonstrate superior technical and business value, technology and tools, customer support, and overall value as evidenced in customer survey scores and interviews with customers. Victors also have significant depth (and often breadth) of functionality and overall strong customer satisfaction

scores. The combined scores across Value and Validity rank at or above the median scores in this category. SAS, OpenText, IBM and Clarabridge are the Victors in this category. Additionally, two companies received special mention: Provalis Research for its strong score in the Value category and Megaputer Intelligence for its strong showing in the Validity category. While they did not receive overall Victor status, we felt it important to highlight them along with the Victors.

- **SAS.** SAS is one of the overall Victors in this category. Customers felt that SAS had a very strong offering since purchasing Teragram and adding new feature/functionality. SAS received the highest score for overall customer satisfaction. It also received the highest score in time to value, perhaps due to the fact that it does well integrating with the SAS Enterprise Miner platform and that customers who are already using the product found the text mining capabilities easy to integrate. The ontology and taxonomy capabilities of the product impressed customers, as did its ease of integration and the overall benefits that the software provides.
- **OpenText.** OpenText is also an overall Victor in this category because it received very high marks across the board. Customers liked the value for the price and the fact that the company is rolling out new feature/functionality. Additionally, customers appreciated the performance of the platform and felt that the solution was quite strong technically.
- **IBM.** Both IBM SPSS and IBM Content Analytics with Enterprise Search received the highest scores against all of the companies profiled for overall business value and overall business benefits. IBM SPSS received very high marks for value/price and its integration capabilities (i.e. as part of its predictive analytics capability). IBM Content Analytics with Enterprise Search received very high marks for support and for some of its tools and technologies, such as its visualization capabilities.
- **Clarabridge.** Clarabridge received some of the highest scores for its feature/functionality roll out and its sentiment analysis capabilities. Customers felt that they were getting a lot of technical benefit from the product set and liked the fact that Clarabridge is very responsive to their needs. They especially liked its eleven-point sentiment scale, which is very useful in voice of the customer analysis. While some end-users thought the software was a bit pricey, they felt its time to value was high.

#### Special Mention- Value

- **Provalis Research:** Customers believe that the tools offered by Provalis Research are extremely well priced. Provalis received the highest scores for price, ROI, and value for the price. While it is not a full-featured product like some of the other vendors profiled, Provalis can help companies tackle a lot of text analytics problems at a very reasonable price point. Its overall value score was quite high compared to the other vendors profiled.

#### Special Mention- Validity

- **Megaputer Intelligence:** Customers liked the feature/functionality of Megaputer and the fact that it can work across both structured and unstructured data. End-users also liked how it handles taxonomy and ontology building. Megaputer also received one of the highest scores for breadth and depth of functionality. Its overall validity score was quite high compared to the other vendors profiled.

*OpenText is also an overall Victor in this category because it received very high marks across the board.*



## Leaders

Leaders have a solid product as indicated by customer survey scores. Customers are generally satisfied with their solutions and believe that they provide good value. However, the scores were not as high as the companies that achieved Victor status. The summary for each Leader below highlights the areas where the company scored well and points out one or two areas where customers would like to see improvement.

- **Attensity:** While customers liked Attensity's tools and thought they provided a good benefit overall, survey results indicate that customers felt that Attensity 5.0 was not all that easy to use and it was a bit pricey. They also felt that the ontology and taxonomy support were lacking. However, they believe that Attensity 6.0 will be a solid leap forward in terms of ease of use and feature/functionality, so we expect that Attensity will move up to the Victor category in the near term.
- **TEMIS:** Customers liked the business value that the TEMIS solution provided and the company received high scores for its ontology and taxonomy support. However, some customers felt that the breadth and depth of the tool set was lacking and that it was not that easy to use. Additionally, they felt it was a bit expensive for the value.
- **Megaputer Intelligence:** Customers liked the breadth and depth of functionality that Megaputer Intelligence provides and particularly its ability to deal with structured and unstructured data. It received a special mention in terms of validity. However, customers would like to see better support for language and would like the company to roll out feature/functionality more quickly.
- **Provalis:** In keeping with its focus to provide sound value at a reasonable price, Provalis received a special mention in terms of value. However, while customers felt that Provalis could provide value in a number of problem areas, survey results indicate that customers felt that sentiment analysis and some other technical capabilities, such as better integration, would help.

## Challengers

Challengers are viewed as potential Leaders or Victors, but are either too new to the market to obtain a good understanding of their strength or there was not enough information to place them in one of the other categories.

- **Basis Technologies:** Basis Technologies provides a strong solution for companies that need multilingual support. However, customers did not provide enough information to formally place the company into the Victor or Leader category from a customer perspective.
- **Lexalytics:** Customers like the price point of Lexalytics, and felt it was easy to use and that its sentiment engine was fairly accurate. However, customers did not provide enough information to formally place the company into the Victor or Leader category. We expect that if we had received more information Lexalytics might have been a Victor, from the customer perspective.
- **Verint:** Verint seemed to receive good marks across the board. However, customers did not provide enough information to formally place the company into the Victor or Leader category.

*Leaders have a solid product as indicated by customer survey scores. Customers are generally satisfied with their solutions and believe that they provide good value.*



- Other companies that scored well in online customer surveys including Alchemy API and Anderson Analytic's Odin Text.

## Vendor Assessment

### SAS

Company: SAS  
Private

Website: [www.sas.com](http://www.sas.com)

Year Founded: 1976

Business Value: With depth and breadth of solutions, SAS supports customers with increasing volumes of complex and unstructured data across multiple problem areas.



*SAS offers a range of products in the text analytics space targeted to users across the analysis spectrum.*

SAS wants to be the first company that customers turn to when they need to solve analytical business problems. When SAS purchased text analytics vendor Teragram in 2008, it wanted to enhance its strategy to use both structured and unstructured data in analysis and to integrate it for descriptive and predictive modeling. It has come a long way since the purchase. Now, its text analytics capabilities are part of its overall analytics platform and text data is viewed as simply another source of data.

The company's overall strategy is presently three fold. First, SAS will continue to innovate in the area of high-performance analytics in order to insure that performance meets customer expectations. The goal is to take problems that used to take weeks and solve them in days, or problems that used to take days and solve them in minutes. Second, SAS is looking to operationalize analytics. It believes that in order to take full advantage of analytics, the results need to be encapsulated in linguistic and business rules, and operational systems. SAS will continue to evolve its technology to insure that all analytics (both structured, unstructured and combinations) can be consumed across the organization in order to drive business decisions. Finally, SAS will continue to push the envelope by delivering new text analytics methodologies in support of the increasingly complex business problems presented by customers.

SAS offers a range of products in the text analytics space targeted to users across the analysis spectrum. Customers like the fact that these products can integrate fairly easily into the SAS platform and that SAS tools are available for analysis of unstructured (and structured) data. SAS also offers numerous vertical solutions and also delivers industry-focused and sentiment taxonomies to help users jumpstart their text analytics work. Additionally, it has recently enhanced its sentiment analysis product by adding mood state taxonomies and sentiment taxonomies.

Customers who already use SAS Enterprise Miner found its Text Miner product easy to get started with, in terms of both ease of use and ease of integration. One customer noted, "It is one platform that is very easy to plug it in to." Even without already using Enterprise Miner, customers noted that, "with only a brief introduction to the tool, anyone analyzing text can use it." Other customers liked the fact that end-users could deal with and analyze structured and unstructured data together. As one customer stated, "At the end of the day, you don't have to be a SAS guru to use the product(s). It's pretty easy to do subsetting using structured and unstructured data together." The SAS feature set is also robust. A number of customers said they were just scratching the surface with "powerful" features. One customer noted that SAS is, "full of tools and very flexible. The product set is evolving and getting better and better." Another customer noted that the sentiment analysis was "unbelievable" and that SAS was a "workhorse" with regard to performance.

*... customers noted that, "with only a brief introduction to the tool, anyone analyzing text can use it."*

When asked, "Who would you recommend this for?" there were a range of answers which reflect the fact that the product suite can be used by a range of end-users for a range of problems.

## SAS Overview

### Products

SAS offers four main text analytics products: SAS Enterprise Content Categorization, SAS Text Miner, SAS Sentiment Analysis, and SAS Ontology Management. These are described below along with other products and solutions SAS offers in the text analytics space.

- **SAS® Enterprise Content Categorization v12.1:** designed to classify and categorize large volumes of multi-lingual content through managed taxonomies and advanced linguistic rules that parse, identify and extract entities, facts and events.
  - SAS® Document Duplication Detection: This add-on technology is designed to identify similar documents based on configurable definitions of similarity; helping organizations reduce content duplication.
  - SAS® Industry Taxonomy Rules: These add-on technologies are designed to reduce initial taxonomy development time with pre-defined parent/child terms and attributes specific to individual industries.
  - SAS® Search and Indexing: This add-on technology is a thin client interface to category facets that supports accurate and linguistic-based search and retrieval queries.
  - SAS® Text Summarization: This add-on technology distils documents down to the most important sentence concepts which is helpful in search activities as well as information delivery to SMS devices.
  - SAS® Web Crawler: This add-on technology downloads materials from the internet, as well as those from internal file systems, and can include content categorization rules



within the extraction criteria to bring the desired materials into analysis and storage systems.

- **SAS® Content Alerts:** This add-on technology is designed to send notifications based on defined criteria to email or SMS systems in a variety of optional formats and is used to notify individuals, update reports or systems to changes in content.
- **SAS® Text Data Language Pack:** While SAS Enterprise Content Categorization includes the English language and the non-English native language (as applicable), organizations can also add-on additional languages to be analyzed with the company's native dictionary-based language packs.
- **SAS® Information Workbench:** This add-on technology is a workflow tool that incorporates automatic abstracting, categorization and entity extraction. It is designed for reducing workloads of indexers or editors.
- **SAS® Content Categorization:** This is a single-user version of SAS Enterprise Content Categorization – defined for SMB/SME organizations that want to build, manage and deploy taxonomies for classifying content. All add-ons available to SAS Enterprise Content Categorization are also available to SAS Content Categorization customers.
- **SAS® Text Miner 12.1:** Designed to discover clusters of information and topics buried in content collections. It also automatically generates predictive rules to classify content. The following are add-ons to the product:
  - **SAS® Concept Creation for SAS Text Miner:** Designed to build, manage and deploy custom entities into SAS Text Miner discovery analysis, allowing the inclusion of pre-defined concepts and entities within the automated discovery of topics.
  - **SAS® Text Miner Language Pack:** While SAS Text Miner includes the English language, and the non-English native language. Organizations can also add additional languages to their text mining analysis with these native, dictionary-based language packs.
- **SAS® Text Miner for Desktop:** Designed for SMB/SME organizations that would like to mine document collections with a single user desktop application.
- **SAS® Enterprise Miner:** While a required product for SAS Text Miner, Enterprise Miner analysis can be used in conjunction with text mining so that predictive and descriptive models are based on both structured and unstructured data.
  - **Credit Scoring for Enterprise Miner:** As an add-on to the Enterprise Miner environment, unstructured text results can be included in credit scorecard development.
  - Unstructured data can also be incorporated into a series of

*SAS Text Miner and Enterprise Miner analysis can be used in conjunction with text mining so that predictive and descriptive models are based on both structured and unstructured data.*



SAS products including SAS/ETS (Econometrics and Time Series, SAS/STAT, SAS/IML, SAS Forecast Server, SAS Operations Research, and SAS JMP PRO

- **SAS® Sentiment Analysis 12.1:** This product is designed to locate and analyze text content in real time in order to hone in on the sentiment. It includes pre-defined templates for common social media sites Google, Facebook, Twitter, Bing, BoardReader, Flickr, LinkedIn and Yahoo! as well as a new GUI interface for point-and click specification for defining fields from specific sites.
  - SAS® Text Data Language Packs: Same add-on as above
- **SAS® Ontology Management 12.1:** This product is designed to centrally manage semantic terms, and their conditional use, in a centralized environment so that siloed content can be semantically associated. This allows organizations to automate related content in retrieval activities.

Target user/  
company size

- Target company size: SAS Text Analytics products are targeted to organizations of all sizes, including small, mid-market and large commercial, government and academic organizations. Deployments exist across all of these segments.
- Target users: Range in function from modelers with a linguistics or text analytics background to subject matter/ business analyst or business manager, who use a point and click interface to explore scored content and provide feedback to the model development for refinement to business users, who interact with the scored content results in dashboards, email alerts, and mobile reports and IT who deals with integration issues.

Verticals  
supported

- While it supports virtually every industry, SAS provides packaged solutions in the following verticals: Financial services, government, healthcare, insurance, manufacturing, media, policing, and retail
- These solutions include SAS Enterprise Financial Crimes, SAS Framework for Insurance, Government, Healthcare, Supply Chain, SAS Criminal Justice and Public Safety, SAS Fusion Centers and Public Security Solutions, SAS Early Warning Detection (government), Citizen Intelligence, Content Analytics (media), and Adverse Event Detection for Public Safety.

Horizontal  
solutions

SAS Social Media Analytics (available as SaaS only), SAS Fraud Framework, Enterprise Case Management, Warranty analysis, Predictive Asset Maintenance, Personal Identifiable Information, Master Data Management and Data Quality.

How offered

On-premises or hosted, with the exception of SAS Social Media Analytics (which is hosted only)

*While it supports virtually every industry, SAS provides packaged solutions in the following verticals: Financial services, government, healthcare, insurance, manufacturing, media, policing, and retail.*



Differentiator(s)	<ul style="list-style-type: none"> <li>• Breadth of technology and tools in an integrated platform</li> <li>• An actual ontology management product</li> <li>• The ability to analyze structured and unstructured information together</li> <li>• Sophisticated analytics for unstructured data</li> </ul>
Platforms supported	<ul style="list-style-type: none"> <li>• HP-UX Itanium, HP-UX 11iv3 (11.31), HP-UX PA-RISC</li> <li>• IBM AIX: Version 6.1 and 7.1 on POWER architectures</li> <li>• Linux for x64, EM64T/AMD64): RHEL 5 and 6, SuSE SLES 10 and 11</li> <li>• Microsoft Windows (x86-64): Windows XP Professional, Windows Vista, Windows 7, Windows Server 2003 family, Windows Server 2008 family</li> <li>• Microsoft Windows on x64 (EM64T/AMD64): Windows XP Professional for x64, Windows Vista for x64, Windows 7 for x64, Windows Server 2003 for x64, Windows Server 2008 for x64</li> <li>• Solaris on SPARC: Version 10</li> <li>• Solaris on x64: Version 10</li> </ul> <p>Supported Web browsers include:</p> <ul style="list-style-type: none"> <li>• Internet Explorer 7 and 8 on Windows XP Pro, Windows Vista and Windows 7</li> <li>• Firefox 3.6 on Windows XP Pro, Windows Vista, Windows 7 and Linux 32-bit, Linux x64</li> </ul> <p>Note: Middle tier required/optional software for SAS Text Miner (only)</p> <ul style="list-style-type: none"> <li>• SAS client and middle tier require Sun JRE 1.6 And SAS® Text Miner for Desktop (only) specific to:</li> <li>• Microsoft Windows (x86-64): Windows XP Professional, Windows Vista, Windows 7</li> <li>• Microsoft Windows(x64): Windows XP Professional for x64, Windows Vista for x64, Windows 7 for x64</li> </ul>
Languages supported natively	<p>SAS Text Analytics (language dependent products) natively supports Arabic, Chinese (both Simplified and Traditional), Czech, Danish, Dutch, English (US/UK), Finnish, French (French/Canadian), German (New/Old), Greek, Hebrew, Hungarian, Indonesian, Italian, Japanese, Korean, Norwegian (Nynorsk/Bokmål), Polish and Portuguese (Portugal/Brazil), Romanian, Russian, Slovak, Spanish (Spain/South America), Swedish, Thai, Turkish and Vietnamese. SAS can identify various languages in a document and using the appropriate dictionary can process each language in a single text/document by running multiple language models against the same content.</p>



## Key Features

- Text analytics techniques used: SAS uses a combination of statistical and NLP approaches, which can be used in conjunction or in isolation, to improve classification and concept extraction as well as discover new topics and terms.
- Supervised, semi-supervised and unsupervised techniques: SAS provides a number of unsupervised learning techniques including: LSA, Naïve Bayesian, and Hierarchical Bayesian methods. In addition, taxonomies and advanced linguistic rules can be used to include domain or business knowledge. In addition, the option is available for results from such methods to be included into unsupervised analysis. The company has recently released a new semi-supervised method for automatic rule generation that supports active learning techniques .
- Extracts: entities, categories, concepts, sentiment, mood and facts. The methodology is also applicable to other scaled measures, for example, expertise or socio-economic status. SAS Text Summarization can summarize documents and SAS Text Miner can discover themes and topics.
- Sentiment analysis: SAS provides a separate product for sentiment analysis (see above). SAS identifies sentiment expressed in a document as well as at detailed feature, attribute and sub-attribute levels. The company states that accuracy has been as high as 98% but this level of accuracy changes per project.
- Support for structured data: SAS can extract non-linguistic entities from text, such as telephone numbers, SSNs/SINs and any others, using regular expressions and/or pre-built operators. Additionally, SAS integrates structured and unstructured data together within the text analytics environment, so structured fields with desired values are retained with the unstructured analysis – both being available to predictive analysis. SAS can also structure unstructured text for direct inclusion into predictive analysis. SAS also offers a suite of tools such as Basic SAS coding, SAS Data Integration and SAS Business Intelligence, which can take the output from SAS Text Analytics and combine it with structured data.
- Document level analysis: SAS can both analyze documents at the individual level as well as an entire collection. By combining statistical and linguistic techniques, SAS is able to analyze and extract information at the document level and also discover concepts and topics from assessments of the entire collection.
- Ontology and taxonomy support: SAS Ontology management manages thesauri and multiple taxonomies and semantic terms. Administration for SAS Ontology Management allows for a range of personnel to apply their domain expertise to the definition and configuration of the system options. Security and user authorization is centrally managed. SAS Enterprise Content Categorization is used to define and manage single taxonomies (in a collaborative environment) and is directly integrated into SAS Ontology Manager. SAS Ontology Manager can import and export OWL and RDF format.
- Integrated text analytics: SAS Text Miner can accept custom entities from categorization, automatically discover Boolean rules for linguistic rule

*SAS can extract non-linguistic entities from text, such as telephone numbers, SSNs/SINs and any others, using regular expressions and/or pre-built operators.*



development for categorization, and create document level training corpus for sentiment analyses. Categorization taxonomies can be used in either sentiment analysis or ontology management. And ontology classifiers and concepts can be used for real-time semantically meaningful classification with categorization.

- Integration with content management systems: SAS can integrate with SharePoint and EMC Documentum.
- Integration with other systems: SAS can integrate with Enterprise Applications including Oracle Enterprise Business Suite, SAP, Siebel, and Salesforce.com. It also integrates with Endeca (Oracle) and Fast (Microsoft).
- Integration with data appliances: SAS Text Analytics integrates with Teradata and EMC/Greenplum. With its 12.1 release it also made high performance text mining available. In addition, SAS can run in-database with many other vendors such as Netezza and Aster Data.
- Other: SAS has multiple techniques to handle “big data”, including:
  - In-stream processing. This scores the data as it comes in or loads. This method processes the data before it is stored reducing any batch processing time.
  - In-database technology. This natively executes relevant SAS software functions inside databases and includes scoring acceleration, analytics acceleration, and in-database analytics, the former two are currently specific to Teradata only.
  - Grid Computing. A centrally managed grid infrastructure that enables parallel processing for data management, analytics and reporting.
  - In Memory processing. This puts all of the data and processing in-memory eliminating I/O restrictions and enabling big data visualizations and analytic modeling functions.

SAS also has a number of access engines to natively read/write data, including SAS/Access to Hadoop which, in addition to the read/write to HDFS capability, can also use other Hadoop capabilities such as Pig and Hive languages and MapReduce

*SAS can integrate with Enterprise Applications including Oracle Enterprise Business Suite, SAP, Siebel, and Salesforce.com. It also integrates with Endeca (Oracle) and Fast (Microsoft).*



## About Hurwitz & Associates

Hurwitz & Associates is a consulting, market research and analyst firm that focuses on how technology solutions solve real world business problems. The firm's research concentrates on disruptive technologies, such as Cloud Computing, Service Oriented Architecture and Web 2.0, Service Management, Information Management, and Social and Collaborative Computing. We help our customers understand how these technologies are reshaping the market and how they can apply them to meet business objectives. The team provides direct customer research, competitive analysis, actionable strategic advice, and thought leadership. Additional information on Hurwitz & Associates can be found at [www.hurwitz.com](http://www.hurwitz.com).



© Copyright 2013, Hurwitz & Associates

All rights reserved. No part of this publication may be reproduced or stored in a retrieval system or transmitted in any form or by any means, without the prior written permission of the copyright holder. Hurwitz & Associates is the sole copyright owner of this publication. All trademarks herein are the property of their respective owners.

This report is based on independent research and input from a random sampling of users. This report represents our findings, some user experiences may vary from our conclusions. Hurwitz & Associates did not receive financial or other benefits in connection with the Victory Index while preparing the report. The information contained in this report has been obtained from sources believed to be reliable. Hurwitz & Associates shall not be liable for any omissions or errors in the information contained in this report. All opinions in this report are subject to change without notice.

175 Highland Avenue, 3rd Floor • Needham, MA 02494 • Tel: 617-597-1724  
www.hurwitz.com