

The Forrester Wave™: Big Data Text Analytics Platforms, Q2 2016

The 10 Providers That Matter Most And How They Stack Up

by Boris Evelson

May 9, 2016

Why Read This Report

In our 20-criteria evaluation of big data text analytics platforms providers, we identified the 10 most significant ones — Attivio, Cambridge Semantics, Clarabridge, Digital Reasoning, Expert System, HP Enterprise (HPE), IBM, Linguamatics, OpenText, and SAS — and researched, analyzed, and scored them. This report shows how each provider measures up and helps application development and delivery (AD&D) professionals working on text analytics and other advanced analytics and big data initiatives make the right choice.

Key Takeaways

IBM, Clarabridge, SAS, Digital Reasoning, HPE, And Attivio Lead The Pack

Forrester's research uncovered a market in which IBM, Clarabridge, SAS, Digital Reasoning, HPE, and Attivio lead the pack. Cambridge Semantics, OpenText, Expert System, and Linguamatics offer competitive options.

Enrich Customer Insights With Unstructured Data

Today, unstructured data is largely untapped for customer insights, and AD&D professionals see this as the next opportunity to further enhance enterprise insight from structured data available inside and outside the enterprise.

Breadth Of The Platform And Ease Of Use Are Key Differentiators

Just having a natural language processing (NLP) engine is not enough for a broad text analytics platform. Vendors that provide a breadth of features, from data ingestion to analysis in an intuitive, user-friendly platform, will dominate the market.

The Forrester Wave™: Big Data Text Analytics Platforms, Q2 2016

The 10 Providers That Matter Most And How They Stack Up



by [Boris Evelson](#)

with [Srividya Sridharan](#), [Maxie Schmidt-Subramanian](#), [Samantha Ngo](#), [Art Schoeller](#), [Cheryl McKinnon](#), [Rowan Curran](#), [Mike Gualtieri](#), [Ian Jacobs](#), Elizabeth Cullen, and Tyler Thurston

May 9, 2016

Table Of Contents

2 Enrich Customer Insights With Unstructured Data

First, Get Your Text Analysis Terminology Straight

Know Your Use Case Before Navigating The Complex Text Analytics Vendor Landscape

5 Big Data Text Analytics Evaluation Overview

Evaluated Vendors And Inclusion Criteria

7 Vendor Profiles

Leaders Offer Broad Text Mining And Text Analytics Functionality

Strong Performers Are Hot On The Leaders' Heels With Highly Competitive Offerings

14 Supplemental Material

Notes & Resources

Forrester conducted product evaluations in February 2016 and interviewed 10 vendor companies: Attivio, Cambridge Semantics, Clarabridge, Digital Reasoning, Expert System, HP Enterprise (HPE), IBM, Linguamatics, OpenText, and SAS.

Related Research Documents

[Market Overview: Text Analytics](#)

[TechRadar™: Customer Analytics Methods, Q2 2016](#)

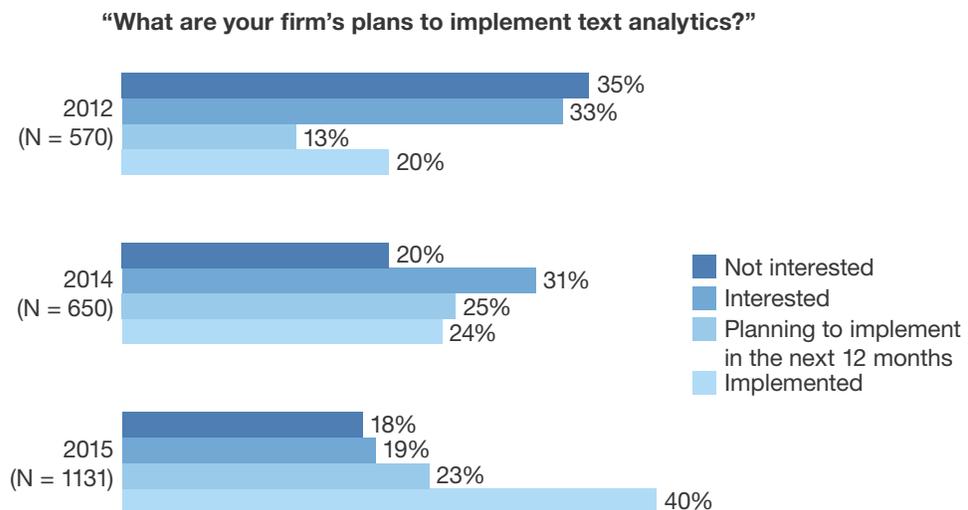
[Vendor Landscape: Big Data Text Analytics](#)

The Forrester Wave™: Big Data Text Analytics Platforms, Q2 2016

The 10 Providers That Matter Most And How They Stack Up

Enrich Customer Insights With Unstructured Data

Over the past three years, Forrester's research has written extensively about the age of the customer. Forrester believes that only the enterprises that are obsessed with winning, serving, and retaining customers will thrive in this highly competitive, customer-centric economy.¹ But in order to get a full view of customer behavior, sentiment, emotion, and intentions, AD&D professionals must help enterprises leverage all the data at their disposal, structured and unstructured.² Alas, that's still an elusive goal, as most enterprises leverage only 40% of structured data and 31% of unstructured data for business and customer insights and decision-making.³ But there is hope: More organizations have implemented text analytics, while fewer organizations report that they are not considering text analytics (see Figure 1).

FIGURE 1 Text Analytics Adoption Trends

Base: 570 to 1,131 North American and European data and analytics decision-makers

Note: “Don't know” responses omitted from analysis.

Source: Forrester's Global Business Technographics® Data And Analytics Survey, 2015; Forrester's Business Technographics Global Data And Analytics Survey, 2014; and Forrester's Forrsights Strategy Spotlight: Business Intelligence And Big Data, Q4 2012

First, Get Your Text Analysis Terminology Straight

For AD&D pros, the process of text mining and text analytics should not be a black box, where unstructured text goes in and structured information comes out. But today, there is a lot of market confusion on the terminology and process of text analytics. The market, both vendors and users,

The Forrester Wave™: Big Data Text Analytics Platforms, Q2 2016

The 10 Providers That Matter Most And How They Stack Up

often uses the terms text mining and text analytics interchangeably; Forrester makes a distinction and recommends that AD&D pros working on text mining/text analytics initiatives adopt the following terminology:

- › **Text mining extracts structures from unstructured text.** Text mining is a complex process where there is a specific sequence of steps to follow. Text mining first involves connecting to data sources, ingesting the text, cleaning it, preprocessing it, and mining it to extract structures like entities, concepts, and sentiment scores.⁴ Most often the process requires several iterations for data enrichment, AKA “training.” Business domain subject matter experts and professional linguists can reiteratively train the system to be more accurate with each iteration by letting the system leverage industry or business domain specific ontologies, taxonomies, and lexicons.
- › **Text analytics analyzes the findings of the text mining process.** Text analytics then answers the question “What have we found?” It’s a process that, via a graphical user interface (GUI), lets the user analyze and organize the findings. Text analytics features include displaying counts of the structures uncovered in the text mining process, their relationships (via network or graph diagrams), and hierarchies (industry, product hierarchies).
- › **Post-processing text analysis uncovers patterns.** Post-processing text analysis analyzes patterns based on various attributes. This step of the process is used to explore time (as in “Is the sentiment trending up or down over time?”), region, customer segment (as in “How does the sentiment vary by customer segment?”) and other patterns by any of the available attributes. In the world of structured data analysis, this is often referred to as online analytical processing (OLAP). In the world of search, this is often referred to as faceted navigation.

Know Your Use Case Before Navigating The Complex Text Analytics Vendor Landscape

To address all of the text mining and text analytics requirements, vendors compete in a broad, diverse, and complex landscape that includes more than 200 potential players. But before considering investing in a specialized text mining and text analytics product, Forrester recommends that AD&D pros investigate whether some of their existing enterprise software platforms, tools, and applications (extract, transform, load [ETL], database management systems [DBMS] and others) already have all, most, or some of the required text mining and/or text analytics features. To help AD&D pros working on text analytics, other advanced analytics, and big data initiatives consider an existing enterprise software platform that may already have text mining and/or text analytics capabilities at Forrester suggests using the following vendor landscape based on use cases and buyer type (see Figure 2).

The Forrester Wave™: Big Data Text Analytics Platforms, Q2 2016

The 10 Providers That Matter Most And How They Stack Up

FIGURE 2 Technologies With Text Mining Or Text Analytics Capabilities

Category	Text analytics capabilities	Use case	Buyer	Example vendors and products
Analytical DBMS	Broad text mining	Improving performance of high number of records processing using scalable DBMS engines.	Tech management	IBM DB2, HPE Vertica, Teradata Aster Analytics, Oracle (via OEM of Lexalytics), SAP Hana, and SAP IQ
Contact center platforms	Mostly speech-to-text mining and analytics	Improving contact center agent quality.	Business	CallMiner and Genesys
Data integration, including extract, load, and transfer	Broad text-mining capabilities	Preparing data for loading into data warehouses, datamarts, or to be used by apps.	Tech management	Ab Initio, Informatica, and SAP
eDiscovery and document classification	Focused on document classification	Classifying documents by categories (e.g., risk, compliance, etc.).	Business	Active Navigation, Exterro, FTI Consulting, HPE Records Manager, IBM StoredIQ eDiscovery, Knowliah, Megaputer, Nuix, and OpenText Discovery
Enterprise content management (ECM)	Focused on document classification	Classifying documents by categories (e.g., risk, compliance, etc.).	Business	IBM, Lexmark Enterprise Software, OpenText
Natural language processing (NLP) APIs	Natural language processing	Embedding NLP into other platforms and customer apps.	Tech management	AlchemyAPI, Apache OpenNLP, Basis Technology, Content Analyst, Google Cloud Machine Learning and HPE Haven OnDemand
Predictive analytics	Broad text-mining capabilities	Leveraging results of text mining into predictive models.	Business	Alpine Data Labs, Alteryx, Angoss Software, Dell, FICO, IBM, KNIME, Microsoft, Oracle, RapidMiner, SAP, SAS, and Teradata Aster Analytics
Search	Broad text-mining capabilities	Improving keyword search results with semantic search and knowledge discovery.	Tech management	AddStructure, Attivio, Coveo Solutions, Google, HPE, IBM, Lexmark International, Lucidworks, Mindbreeze, Oracle BigData Discovery, and Squirro
Social media listening	Focused on sentiment analysis	Garnering insights from social and mainstream media	Business	Brandwatch, Clarabridge, EPAM, Infegy, NetBase, Synthesio, and Sysomos
Voice of the customer	Focused on sentiment analysis	Garnering insights from customer feedback management	Business	Ascribe, Clarabridge, Confirmat, Feedback Ferret, IBM, InMoment, MaritzCX Research, Medallia, NICE Systems, Qualtrics, Questback, ReponseTek Networks, Satmetrix Systems, and Taste Analytics

The Forrester Wave™: Big Data Text Analytics Platforms, Q2 2016

The 10 Providers That Matter Most And How They Stack Up

Big Data Text Analytics Evaluation Overview

To assess the state of the big data text analytics market and see how the vendors stack up against each other, Forrester evaluated the strengths and weaknesses of the top big data text analytics platform vendors. After examining past research, user need assessments, and vendor and expert interviews, we developed a comprehensive set of evaluation criteria. We evaluated vendors against 20 criteria, which we grouped into three high-level buckets:

- › **Current offering.** To evaluate a vendor's current offering, we looked at functional and technical capabilities, such as connectivity to a variety of text data sources, text ingestion and preprocessing, text mining based on linguistic and/or statistical rules, data enrichment with business-domain-specific ontologies for improved accuracy, GUI, and architecture and scalability. The 11-evaluation criteria included in the current offering bucket have approximately 200 detailed quantitative evaluation measures rolled up into them, which vendors responded to and Forrester confirmed during product briefings and demos (which included both Forrester- and vendor-supplied data sets).⁵ The results and the scores were also validated by vendor client reference surveys and interviews.
- › **Strategy.** We evaluated and validated the vendor's strategy and vision, intellectual property, and partnership ecosystem. The results and the scores were also validated by vendor client reference surveys and interviews.
- › **Market presence.** We evaluated market presence based on the vendor's installed client base, revenues, and global presence.

Evaluated Vendors And Inclusion Criteria

This Forrester Wave evaluation represents just the tip of the text analytics market. Out of the approximately 200 vendors that Forrester is aware of and/or tracks in this market segment, the 10 vendors Forrester included in this assessment — Attivio, Cambridge Semantics, Clarabridge, Digital Reasoning, Expert System, HPE, IBM, Linguamatics, OpenText, and SAS — already represent the top 5% of the market. Each of these vendors has (see Figure 3):

- › **A complete self-contained text mining and analytics platform.** Each vendor included in this study has a complete self-contained platform. We did not include text analytics components embedded in other applications and platforms. We did not include vendors that only provide text analytics APIs or only some components of a complete platform, such as natural language processing (NLP).
- › **Broad, general-purpose text mining and analytics platform.** Each vendor in this study can support multiple enterprise uses cases across functional areas. We did not include vendors that provide only prebuilt text mining/text analytics applications such as voice-of-the-customer (VoC) or sentiment analysis. We also did not include vendors whose platforms only target specific use cases, like document classification or semantic search.

The Forrester Wave™: Big Data Text Analytics Platforms, Q2 2016

The 10 Providers That Matter Most And How They Stack Up

- › **Market presence as determined by revenues and interest from Forrester clients.** Each vendor generates over \$10 million in annual revenue from text mining/analytics software licenses or subscriptions. Also, Forrester clients often ask about these vendors during relevant inquiry calls.

FIGURE 3 Evaluated Vendors: Product Information And Selection Criteria

Vendor	Product	Version
Attivio	Attivio Content Analytics	5
Cambridge Semantics	Anzo Unstructured	3.5
Clarabridge	Clarabridge Intelligence Platform	7
Digital Reasoning	Synthesys	3.11
Expert System	Cogito Discover	
HP Enterprise	HPE IDOL	11
IBM	IBM Watson Explorer	11
Linguamatics	I2E	4.4
OpenText	OpenText InfoFusion	10.6
SAS	SAS Contextual Analytics	14.1

Vendor selection criteria

A complete self-contained text mining and analytics platform. Each vendor included in this study has a complete self-contained platform. We did not include text analytics components embedded in other applications and platforms. We did not include vendors that only provide text analytics APIs or only some components of a complete platform, such as natural language processing (NLP).

Broad, general-purpose text mining and analytics platform. Each vendor in this study can serve enterprise needs with multiple uses cases across functional areas. We did not include vendors that provide only prebuilt text mining/text analytics applications such as voice -of-the-customer (VoC) or sentiment analysis. We also did not include vendors whose platforms only target specific use cases like document classification or semantic search.

Market presence as determined by revenues and interest from Forrester clients. Each vendor included generates over \$10 million in annual revenue from text mining/analytics software license or subscriptions. Also, Forrester clients often ask about these vendors during relevant inquiry calls.

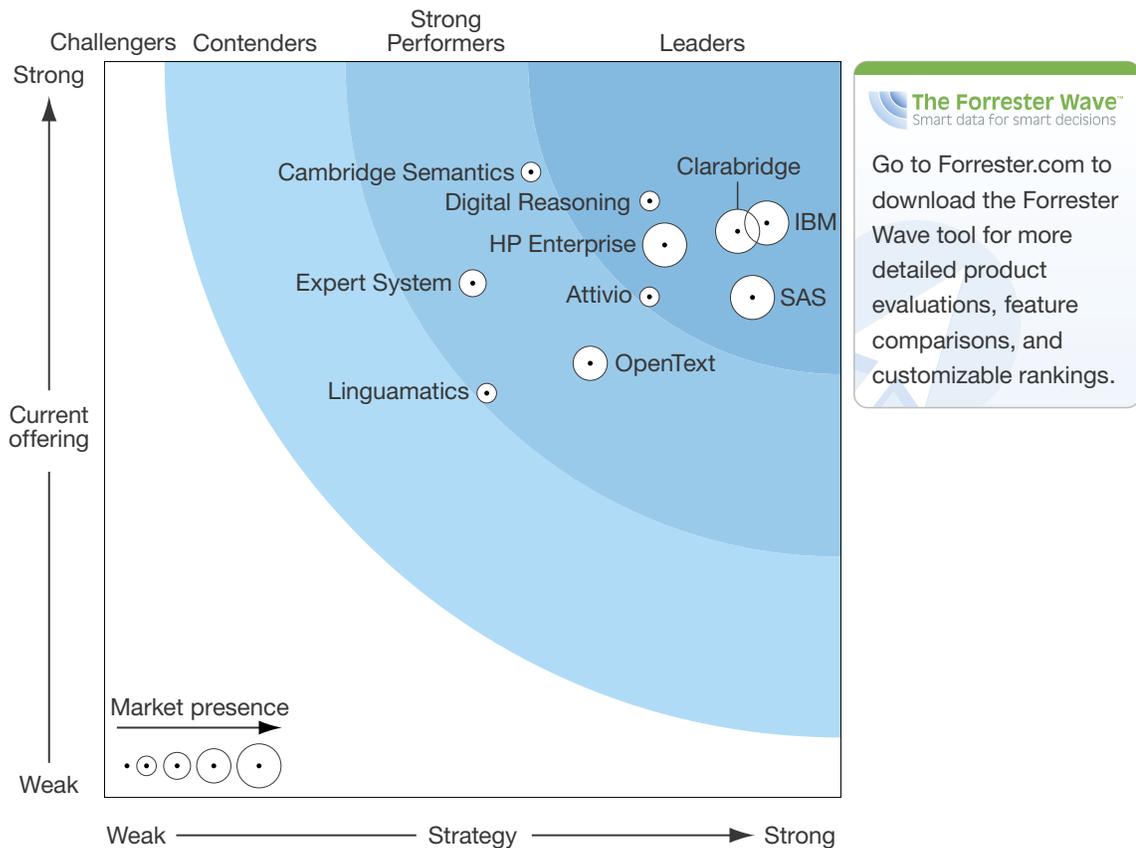
The Forrester Wave™: Big Data Text Analytics Platforms, Q2 2016

The 10 Providers That Matter Most And How They Stack Up

Vendor Profiles

This evaluation of the big data text analytics market is intended to be a starting point only. We encourage clients to view detailed product evaluations and adapt criteria weightings to fit their individual needs through the Forrester Wave Excel-based vendor comparison tool (see Figure 4).

FIGURE 4 Forrester Wave™: Big Data Text Analytics Platforms, Q2 '16



The Forrester Wave™: Big Data Text Analytics Platforms, Q2 2016

The 10 Providers That Matter Most And How They Stack Up

FIGURE 4 Forrester Wave™: Big Data Text Analytics Platforms, Q2 '16 (Cont.)

	Forrester's Weighting	Attivio	Cambridge Semantics	Clarabridge	Digital Reasoning	Expert System	HP Enterprise	IBM	Linguamatics	OpenText	SAS
CURRENT OFFERING	50%	3.40	4.25	3.85	4.05	3.50	3.75	3.90	2.75	2.95	3.40
Document ingestion capabilities	15%	4.00	4.00	3.00	3.00	3.00	4.00	4.00	1.00	2.00	2.00
Text preprocessing capabilities	5%	4.00	5.00	4.00	4.00	4.00	4.00	4.00	2.00	3.00	3.00
Linguistic text preprocessing capabilities	10%	3.00	5.00	5.00	3.00	4.00	3.00	5.00	3.00	4.00	4.00
Advanced linguistic and statistical text mining	10%	3.00	5.00	3.00	5.00	3.00	3.00	3.00	5.00	3.00	3.00
Data enrichment capabilities	10%	3.00	3.00	4.00	4.00	5.00	4.00	3.00	4.00	2.00	3.00
Product accuracy	5%	3.00	3.00	3.00	5.00	3.00	3.00	3.00	5.00	3.00	3.00
User interface (UI)	10%	4.00	5.00	5.00	4.00	4.00	3.00	5.00	2.00	3.00	3.00
Technical architecture, scalability	5%	4.00	4.00	4.00	5.00	4.00	5.00	4.00	4.00	4.00	5.00
Industry and business domain expertise	15%	3.00	5.00	3.00	4.00	3.00	5.00	3.00	1.00	3.00	5.00
Delivery/deployment options	10%	3.00	3.00	5.00	5.00	3.00	3.00	5.00	3.00	3.00	3.00
Customer satisfaction	5%	4.00	4.00	4.00	4.00	3.00	4.00	4.00	4.00	4.00	4.00
STRATEGY	50%	3.70	2.90	4.30	3.70	2.50	3.80	4.50	2.60	3.30	4.40
Key vendor/product differentiators and gaps	50%	4.00	3.00	4.00	4.00	3.00	4.00	4.00	3.00	4.00	4.00
Intellectual property	10%	3.00	1.00	3.00	3.00	1.00	5.00	5.00	1.00	3.00	5.00
Partnership ecosystem	30%	3.00	3.00	5.00	3.00	2.00	3.00	5.00	2.00	2.00	5.00
Cost	0%	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Customer satisfaction	10%	5.00	4.00	5.00	5.00	3.00	4.00	5.00	4.00	4.00	4.00
MARKET PRESENCE	0%	1.40	1.40	4.60	1.40	2.20	5.00	4.60	1.40	3.40	4.20
Installed base	40%	1.00	1.00	5.00	1.00	3.00	5.00	5.00	1.00	3.00	5.00
Revenue	40%	1.00	1.00	5.00	1.00	1.00	5.00	5.00	1.00	3.00	3.00
Global presence	20%	3.00	3.00	3.00	3.00	3.00	5.00	3.00	3.00	5.00	5.00

All scores are based on a scale of 0 (weak) to 5 (strong).

The Forrester Wave™: Big Data Text Analytics Platforms, Q2 2016

The 10 Providers That Matter Most And How They Stack Up

Leaders Offer Broad Text Mining And Text Analytics Functionality

- › **IBM stands up to its brand, adding text analytics cognitive capabilities.** IBM Watson Explorer's architecture allows it to address a use case not all other vendors can offer. Watson Explorer (originally acquired from Vivisimo in 2012) can seamlessly ingest and process structured and/or unstructured data, and perform search or analytical queries. It is also one of the few platforms evaluated in this Forrester Wave that is continuously moving more components into Hadoop, taking advantage of the low-cost distributed processing. This approach is attractive to the clients that are looking to reduce the footprint of software outside of Hadoop clusters. And IBM Watson Explorer goes beyond linguistic and statistical rules — it enriches the capabilities with over 12 cognitive APIs (built by the Watson team and acquired from AlchemyAPI in 2015). The Weather Company and other similar acquisitions also help IBM build its domain-specific ontologies.⁶

Customer references for the most part positively rated IBM's text mining/text analytics platform. However, Watson Explorer is a platform, not a packaged application. While Watson Explorer has preconfigured capabilities that can deliver out-of-the-box value, clients will have to work with IBM to configure and tune it to specific use cases. Luckily, IBM has a large professional services organization with the global reach, experience, and expertise to help clients with all big data and analytics applications. And even though IBM is constantly working to map its enormously diverse software portfolio to specific use cases and streamline the sales process, clients should be prepared to navigate a complex landscape of before zeroing in on the right product. Therefore, Forrester recommends IBM Global Business Services as the best point of entry for customers into IBM text mining (and all other analytics) products.

- › **Clarabridge should be on your enterprise shortlist for VoC applications.** Clarabridge is a full-featured text mining/text analytics platform, with its main strength and strategic direction revolving around VoC and voice-of-the-employee (VoE) applications. One of the key differentiators is its plug-and-play setup, resulting in faster time to deployment. This benefit comes from two key Clarabridge strengths: prebuilt connectors to most VoC and VoE sources (like social media and survey platforms), as well as out-of-the-box domain specific ontologies, taxonomies, and lexicons. Additionally, while most text mining/text analytics platforms provide only basic post-processing analysis features and rely on third-party BI products, Clarabridge's data visualization and slice-and-dice capabilities are almost on par with leading BI platforms.

Customer references viewed Clarabridge mostly positively and especially appreciated frequent product updates. But while the engine behind the Clarabridge platform is an all-purpose text mining/text analytics platform, customers should not consider Clarabridge for non-VoC or non-VoE use cases such as document classification, fraud management, legal eDiscovery, and other multiple use cases. Therefore, consider Clarabridge as an enterprise VoC and VoE platform, but not a broad text mining/text analytics platform for other enterprise applications.

The Forrester Wave™: Big Data Text Analytics Platforms, Q2 2016

The 10 Providers That Matter Most And How They Stack Up

- › **SAS Contextual Analysis seamlessly integrates with the rest of SAS platforms.** SAS is a huge (>\$3 billion in 2015 revenues and uninterrupted growth for 39 years), privately held company, that grew mostly organically, with very few acquisitions. As a result, most of SAS information management technologies, including SAS Visual Analytics and SAS Data Management, are seamlessly integrated and reuse numerous components. SAS also leverages its expertise in text mining/text analytics to build on its strength as a Leader in “The Forrester Wave: Big Data Predictive Analytics Solutions, Q2 2015.”⁷ By combining linguistic parsing with machine learning for document categorization and sentiment, SAS Contextual Analysis is able to achieve better classification with disambiguation, and entity recognition, than statistical rules alone.

Customer references viewed SAS positively as a text mining/text analytics platform vendor, and highly rated its global presence and availability of local resources. However, Forrester feels that SAS Contextual Analysis could come with more prepackaged connectors to popular data sources, especially social media. It does come packaged with SAS Information Retrieval Studio for accessing data from the web. But those customers seeking to add plug-and-play access to social media will have to rely on add-on products from SAS’ cloud-based social media analytics service (via partner products). SAS also sells a separate SAS Text Miner product (not reviewed in this research), which provides even deeper text analytics, with a particular focus on data mining and predictive modeling. Next year, SAS plans to integrate both products into a single platform.

- › **Digital Reasoning brings the future of cognitive computing into text analytics.** Most of the text mining/text analytics platforms combine linguistic and statistical rules, but Digital Reasoning’s engine is based on cognitive computing technology.⁸ This differentiates Digital Reasoning from most competitors in several ways. First, the platform does not require training (but it can be trained, if necessary) and tuning, which can often result in shorter deployment cycles. Next, the process will most likely uncover structures that you didn’t even know existed, addressing the typical “I-don’t-know-what-I-don’t-know” quagmire of most analytics tools. And finally, Digital Reasoning will find entity and concept relationships across multiple languages, automatically build synonyms, and therefore, constantly improve the results. Digital Reasoning is also one of the few vendors in this Wave evaluation that has significant independent third-party verification of its text mining/text analytics accuracy results.

Customer references viewed Digital Reasoning mostly positively, as well as rated the product’s GUI and ease of use highly. However, experienced clients who have rich ontologies, taxonomies, and lexicons to train and enrich text mining/text analytics applications will find that trained platforms based on linguistic rules are more accurate. Also, since Digital Reasoning is not based on any linguistic rules, it does not support popular features such as spelling and grammar correction. Lastly, while Digital Reasoning provides basic data visualization and post-text-mining analysis features, most of the clients should look for connectors and integration with their preferred data visualization and analysis platforms.

The Forrester Wave™: Big Data Text Analytics Platforms, Q2 2016

The 10 Providers That Matter Most And How They Stack Up

- › **HPE's IDOL extends unstructured data analytics to all rich media data sources.** HPE IDOL (the latest version built on the base of a platform acquired from Autonomy in 2011) is a unified platform for text, video, image, and audio analytics to enable insights, while its modular offerings allow flexible deployments. Connectivity to a variety of data sources is an Achilles heel of many other text mining/text analytics vendors, but IDOL supports out-of-the-box connectivity to around 1,000 file types, and access to more than 150 data repositories with manage-in-place capability. This glues separate data silos together without data relocation risks and unnecessary storage costs.

Customer references viewed HPE as a text mining/text analytics platform vendor mostly positively, and highly rated the return on investment (ROI) on their IDOL investments. IDOL is a platform, not a packaged application; clients will have to work with HPE to configure and tune it to specific use cases — luckily, HPE has a large professional services organization with experience and expertise to help clients with all big data and analytics applications. HPE does not provide industry vertical solutions, so clients looking for the vendor to provide domain-specific ontologies, taxonomies, and lexicons will have to look elsewhere (although IDOL can automatically find and generate taxonomies).

- › **Attivio breaks barriers between structured and unstructured data, search, and BI.** Attivio's architecture — an inverted index database management system (DBMS) — makes it agnostic to the data type it indexes (structured or unstructured). While earlier generation enterprise search platforms did a fine job at finding data (but not analyzing it), and business intelligence (BI) platforms helped with analyzing the data (but not finding it in the first place), Attivio combines the best of both worlds. Attivio can also turn parts of that index into relational structures and plug into other popular data visualization tools for more advanced data visualization and analysis. Attivio's road map calls for moving more components into Hadoop; taking advantage of low-cost distributed processing. This approach is attractive to the clients looking to reduce the footprint of software outside of Hadoop clusters.

Customer references viewed Attivio very positively, but had some reservations about the product's GUI and the firm's help desk. In addition to evaluating Attivio's native NLP capabilities for English-language text, Forrester advises clients to evaluate its third-party NLP technologies for non-English languages since Attivio currently OEMs those capabilities. Attivio is a platform, not a commercial-off-the-shelf (COTS) application, so clients will have to work with Attivio to configure and tune it to specific use cases. Attivio does not provide industry vertical solutions, so clients looking for the vendor to provide domain-specific ontologies, taxonomies, and lexicons (key to training and tuning any text analytics application to achieve high degrees of accuracy) will have to look elsewhere.

Strong Performers Are Hot On The Leaders' Heels With Highly Competitive Offerings

- › **Cambridge Semantics offers an RDF and a multiple NLP-based integrated platform.** Any text-mining technology find triples in text — subject-verb-object or subject-relationship-object. Cambridge Semantics' architecture is based on an RDF-compliant graph DBMS that makes this process more efficient and effective. This architecture also supports standardized querying of the

The Forrester Wave™: Big Data Text Analytics Platforms, Q2 2016

The 10 Providers That Matter Most And How They Stack Up

content via SparQL. Additionally, instead of using one NLP engine (which it can, if that's what your requirements call for), Cambridge Semantic uses multiple (native and OEM'd), matching strengths of each one to specific applications. This approach has two major benefits. First, it often uncovers more structures in text than a single engine. Next, Cambridge Semantics' approach borders on cognitive computing: Every time text mining discovers a new concept, it stores it in a knowledge graph. This newly discovered knowledge can then be applied to other applications based on new data sources.

Customer references highly rated Cambridge Semantics Professional Services organization and the ROI on their investments. If your particular use case calls for multiple NLP engines, Forrester advises clients to also evaluate these third-party technologies. Cambridge Semantics is a platform, not a packaged application; clients will have to work with Cambridge Semantics to configure and tune it to specific use cases. Lastly, while Cambridge Semantics provides some data visualization and post-text-mining analysis features, clients looking for more advanced features may want to consider connectors and integration with their preferred data visualization and analysis platforms.

- › **OpenText goes to market with broad enterprise info management platforms.** OpenText InfoFusion (built on the acquisition of Nstein Technologies in 2010) includes comprehensive connectivity to enterprise content repositories tools by way of connectors and a scalable extract, transform, load (ETL) engine as part of the platform. Similar to a couple of other vendors evaluated in this Forrester Wave, OpenText also offers BI and analytics tools for a well-rounded offering — OpenText Big Data Analytics and OpenText Information Hub (iHub) — based on its 2015 acquisition of Actuate.

Customer references rated OpenText InfoFusion's ease of use very highly, but expressed concerns about the frequency of the product release cycles. Additionally, InfoFusion is a platform, not a packaged application; clients will have to work with OpenText to configure and tune it to specific use cases.

- › **Expert System mines text based on a repository of concepts — a knowledge graph.** How do you know whether the word "jaguar" is a large predatory cat or a luxury car? Or if the word "gas" means gasoline or the car accelerator pedal? Expert System easily disambiguates sentences based on a rich, constantly evolving repository of concepts — no additional system training required. However, to address specific domain applications, the product does come with a programming-language-like editor to train and tune the system. Once the text mining process uncovers the concepts and relationships, the product provides an intuitive, faceted-based search navigation to examine the results. And when you don't know what a term means, just hover your mouse over a keyword and an explanation pops up.

Customer references voiced concerns about the product's ease of use for a non-technical user and therefore only leveraged a small percent of all Expert System's capabilities. Additionally, Expert System is a pure-play text mining/semantic search platform. Therefore, it's not a best fit for

The Forrester Wave™: Big Data Text Analytics Platforms, Q2 2016

The 10 Providers That Matter Most And How They Stack Up

all enterprise scenarios, such as classification of large documents or analysis of the text-mining results (such as time series or other patterns). Clients should consider Expert System in addition to, not instead of an all-purpose enterprise text mining/text analytics platform.

- › **Linguamatics delivers deep text analytics specifically for the healthcare industry.** Even though Linguamatics allows clients to plug in their own ontologies, taxonomies, and lexicons, it provides these out of the box specifically for pharmaceutical, biotech, and healthcare industries. These are included in its access to ready-to-use data sources, which include ClinicalTrials.gov, the FDA Online Label Repository, NIH Grants, OMIM, PubMed Central Open Subset, questions and answers on the FDA's Adverse Event Reporting System (FAERS), and Thomson Reuters Cortellis. One of the unique domain expertise of the product is recognizing chemical structures, which makes the product a perfect fit for drug discovery processes. Linguamatics is also one of the few vendors in this Wave evaluation that has significant independent third-party (mostly healthcare sciences academia) verification of its text mining/text analytics accuracy results.

Clients surveyed in this research viewed Linguamatics very positively as a pharmaceutical, biotech, and healthcare text mining/text analytics vendor. They also highly rated the vendor's professional services organization and help desk. While Linguamatics truly stands out as a differentiated vendor in the pharmaceutical, biotech, and healthcare space, it has little to no experience in other verticals and business domains and therefore cannot be considered as your enterprise's one and only text mining/text analytics platform.

Engage With An Analyst

Gain greater confidence in your decisions by working with Forrester thought leaders to apply our research to your specific business and technology initiatives.

Analyst Inquiry

Ask a question related to our research; a Forrester analyst will help you put it into practice and take the next step. Schedule a 30-minute phone session with the analyst or opt for a response via email.

Learn more about inquiry, including tips for getting the most out of your discussion.

Analyst Advisory

Put research into practice with in-depth analysis of your specific business and technology challenges. Engagements include custom advisory calls, strategy days, workshops, speeches, and webinars.

Learn about interactive advisory sessions and how we can support your initiatives.

The Forrester Wave™: Big Data Text Analytics Platforms, Q2 2016

The 10 Providers That Matter Most And How They Stack Up

Supplemental Material

Online Resource

The online version of Figure 2 is an Excel-based vendor comparison tool that provides detailed product evaluations and customizable rankings.

Data Sources Used In This Forrester Wave

Forrester used a combination of three data sources to assess the strengths and weaknesses of each solution. We evaluated the vendors participating in this Forrester Wave, in part, using materials that they provided to us by March 31, 2016.

- › **Vendor surveys.** Forrester surveyed vendors on their capabilities as they relate to the evaluation criteria. Behind the 20 criteria, there were over 200 detailed questions about product architecture, capabilities and functionalities.
- › **Product demos.** We asked vendors to conduct demonstrations of their products' functionality. We used findings from these product demos to validate details of each vendor's product capabilities.
- › **Customer surveys and reference calls.** To validate product and vendor qualifications, Forrester fielded an online survey with eight of each vendor's current customers. We also conducted reference calls with one of those customers.

The Forrester Wave Methodology

We conduct primary research to develop a list of vendors that meet our criteria to be evaluated in this market. From that initial pool of vendors, we then narrow our final list. We choose these vendors based on: 1) product fit; 2) customer success; and 3) Forrester client demand. We eliminate vendors that have limited customer references and products that don't fit the scope of our evaluation.

After examining past research, user need assessments, and vendor and expert interviews, we develop the initial evaluation criteria. To evaluate the vendors and their products against our set of criteria, we gather details of product qualifications through a combination of lab evaluations, questionnaires, demos, and/or discussions with client references. We send evaluations to the vendors for their review, and we adjust the evaluations to provide the most accurate view of vendor offerings and strategies.

We set default weightings to reflect our analysis of the needs of large user companies — and/or other scenarios as outlined in the Forrester Wave evaluation — and then score the vendors based on a clearly defined scale. We intend these default weightings to serve only as a starting point and encourage readers to adapt the weightings to fit their individual needs through the Excel-based tool. The final scores generate the graphical depiction of the market based on current offering, strategy, and market presence. Forrester intends to update vendor evaluations regularly as product capabilities and vendor strategies evolve. For more information on the methodology that every Forrester Wave follows, go to <http://www.forrester.com/marketing/policies/forrester-wave-methodology.html>.

The Forrester Wave™: Big Data Text Analytics Platforms, Q2 2016

The 10 Providers That Matter Most And How They Stack Up

Integrity Policy

We conduct all our research, including Forrester Wave evaluations, in accordance with our Integrity Policy. For more information, go to <http://www.forrester.com/marketing/policies/integrity-policy.html>.

Endnotes

- ¹ Empowered customers are disrupting every industry — and CIOs need to understand how technology management must adapt in this rapidly evolving world. To understand what define the age of the customer and how CIOs must meet these new dynamics, see the “[Technology Management In The Age Of The Customer](#)” Forrester report.
- ² The big promise of big data to deliver better insights to enterprise decision-makers will not fully materialize unless business intelligence (BI) and analytics incorporate all available sources of information, including unstructured data. To understand the typical components and process flows of text analytics, see the “[Market Overview: Text Analytics](#)” Forrester report.
- ³ These percentages are estimates based on reported ranges; the values are not exact. Source: Forrester’s Global Business Technographics® Data And Analytics Survey, 2015.
- ⁴ Forrester tracks over 150 text analytics providers in a diverse and highly fragmented market. We selected the top 42 vendors for a detailed analysis. Each vendor completed a questionnaire about over 200 capabilities. We segmented the vendors into five categories: application only; platforms/applications; customizable applications; customizable platforms; and general purpose customizable platforms. The report also includes a detail explanation of the text analytics components and process flow. Finally, we outlined which vendors use linguistic versus statistical analysis, as well as supervised versus unsupervised platforms. To learn more a specific platform, see the “[Vendor Landscape: Big Data Text Analytics](#)” Forrester report.
- ⁵ Forrester acknowledges Infegy and Haystac providing input on the evaluation criteria.
- ⁶ First, IBM morphed from “big iron” to software and services. Now, it’s transforming again. When IBM announced the acquisition of The Weather Company, it wasn’t about data; rather insight. To learn more about Forrester’s take on this strategic acquisition, see the “[Quick Take: IBM Forecasts Insights-Driven . . . And Buys The Weather Company](#)” Forrester report.
- ⁷ Predictive analytics is within easy reach for all enterprises if they choose the right big data predictive analytics solution to meet their needs. In Forrester’s 45-criteria evaluation, we identified 13 big data predictive analytics solutions providers — Alpine Data Labs, Alteryx, Angoss Software, Dell, FICO, IBM, KNIME.com, Microsoft, Oracle, Predixion Software, RapidMiner, SAP, and SAS — and researched, analyzed, and scored their current market offerings. To learn how each vendor fared against our criteria, see the “[The Forrester Wave™: Big Data Predictive Analytics Solutions, Q2 2015](#)” Forrester report.
- ⁸ While there are multiple definitions of cognitive computing, for the purposes of this report, Forrester defines it as a combination of machine learning, natural language processing user interface, a knowledge graph of people/places/events, and a self-learning and constantly improving technology all based on neural network and other artificial intelligence algorithms. For a better understanding of ideas behind cognitive engagement, see the “[Cognitive Engagement: A New Force Of Creative Destruction](#)” Forrester report.

We work with business and technology leaders to develop customer-obsessed strategies that drive growth.

PRODUCTS AND SERVICES

- › Core research and tools
- › Data and analytics
- › Peer collaboration
- › Analyst engagement
- › Consulting
- › Events

Forrester's research and insights are tailored to your role and critical business initiatives.

ROLES WE SERVE

Marketing & Strategy Professionals

CMO
B2B Marketing
B2C Marketing
Customer Experience
Customer Insights
eBusiness & Channel Strategy

Technology Management Professionals

CIO
› Application Development & Delivery
Enterprise Architecture
Infrastructure & Operations
Security & Risk
Sourcing & Vendor Management

Technology Industry Professionals

Analyst Relations

CLIENT SUPPORT

For information on hard-copy or electronic reprints, please contact Client Support at +1 866-367-7378, +1 617-613-5730, or clientsupport@forrester.com. We offer quantity discounts and special pricing for academic and nonprofit institutions.