FORRESTER®
CHALLENGE THINKING. LEAD CHANGE.
Expand Your Big Data Capabilities With Unstructured Text Analytics

Boris Evelson, VP, Principal Analyst
# Properly deployed, text analytics (TA) offers tremendous tangible benefits

<table>
<thead>
<tr>
<th>Use case</th>
<th>Client</th>
<th>Benefit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Improving customer experience</td>
<td>Orbitz</td>
<td>Ranking of No. 1 travel website in overall customer satisfaction by the American Customer Satisfaction Index</td>
</tr>
<tr>
<td>Email customer requests. Cost savings and efficiency gains</td>
<td>Telecom</td>
<td>$880K in cost savings</td>
</tr>
<tr>
<td>Document search. Avoiding M&amp;A risk</td>
<td>Pharma</td>
<td>Cancelled deal, saving $100 million</td>
</tr>
<tr>
<td>Early warning systems detecting risks and threats to human health</td>
<td>GPHIN</td>
<td>Early detection of H1N1 geographic pandemic</td>
</tr>
</tbody>
</table>
Properly deployed, TA offers tremendous tangible benefits

<table>
<thead>
<tr>
<th>Use case</th>
<th>Client</th>
<th>Benefit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Resolving quality issues</td>
<td>PC manufacturer</td>
<td>Reduce warranty costs by 10% to 15%. Additionally, achieved 30% reduction in general information calls to the contact center.</td>
</tr>
<tr>
<td>Compliance. Improving customer retention and</td>
<td>Bank</td>
<td>Annual value of $11.8 million. 7% increase in customers’ willingness to recommend the company</td>
</tr>
<tr>
<td>opportunities for cross-sell and upsell</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
TA adoption and plans continue to increase

What are your firm’s plans to use text analytics?

- Planning to implement within the next 12 months: 13% (2015), 14% (2016*)
- Implementing/implemented: 30% (2015), 34% (2016*)
- Expanding/upgrading implementation: 21% (2015), 23% (2016*)

Base: 1805 and *2094 global data and analytics technology decision makers.
Source: Business Technographics® Global Data & Analytics Survey, 2016 and 2015
But we use only a small fraction of enterprise data

But anecdotal evidence and deeper dives often show that less than 10% of unstructured and less than 20% of structured data are only being turned into information.

Base: 2094 global technology decision-makers who know how much analytics data their firm uses.
Note: These percents are estimates based on ranges reported. The values are not exact.
Source: Forrester’s Business Technographics® Global Data And Analytics Survey, 2016
Text Analytics definition

- Text mining extracts structures from unstructured text
- Text analytics analyzes the findings of the text mining process
- Post-processing text analysis uncovers patterns
Multiple challenges at every step of the TA process

› Email
› OCR
› Broken code
› Incorrect grammar
› Multiple languages
› Accuracy vs. breadth
› Domain knowledge
› Long cycles
› Integration with BI
Buy vs Build options

- Analytical DBMS
- Contact center platforms
- ETL platforms
- eDiscovery and document classification platforms
- NLP engines and APIs
- Predictive analytics platforms
- Search platforms
- Social media listening applications
- VOC applications
Who’s who in TA

- Broad platform
- Not embedded
- >$10M revenue

Evaluate 10

Track ~40

Landscape of >150 vendors

Attivio, Cambridge Semantics, Clarabridge, Digital Reasoning, Expert System, HPE, IBM, Linguamatics, OpenText, SAS

TA vendor shortlisting and selection

1. Make sure you know your use case
2. Are you looking for a platform, application, or APIs?
3. Are you looking for a cloud, on premise or a hybrid solutions?
4. Are you the end user of TA or are you embedding TA into your solutions?
5. Does the vendor have your industry and business domain expertise?
6. Do you require, are you looking for TA professional services?
We use 200 evaluation criteria for each of the TA functions such as

- Data source connectivity
- Text ingestion
- Text pre-processing
- NLP
- Statistical text mining
- AI and Cognitive
- Data enrichment
- Text analysis
Which view of big data is most aligned to you?

- It’s an extension of existing analytics and BI practices suited for data that is larger or faster than we are used to: 34%
- It’s a whole new way of thinking about the value in data that requires new analytics and leverages some new technologies: 29%
- It’s about new technologies that allow us to handle more data: 22%
- The term “big data” is very confusing; not sure what it means: 8%
- It’s a bunch of hype with little substance and few new ideas: 7%

Base: 3343 global data and analytics decision makers.
Which of the following definitions of big data is the most helpful in framing your thinking and informing your investment and strategy decisions?

Big data refers practices and technologies that close the gap between all the data that’s available and businesses abilities

Big data is techniques and technology that make managing and analyzing data at extreme scale more affordable.

Big data is data, due to its volume, velocity, variety or variability of format has become too difficult or expensive to manage

None of these

Base: 3343 global data and analytics decision makers.
Which of the following are included in your plans for big data?

Public cloud big data services: 40%
Large scale predictive modeling, data mining or other: 36%
Streaming analytics / computing: 33%
Distributed in memory databases, grids, analytics tools: 30%
Unstructured data mining / analytics: 28%
Packaged analytics technologies that brand themselves: 27%
Marketing or digital data management platforms and: 26%
Creating or building out a data lake: 26%
Data anonymization or de-identification: 23%
Hadoop: 23%
Semantic technologies: 22%
A MPP data warehouse: 18%
NoSQL other than Hadoop: 16%
Don’t know: 8%

Base: 2094 global data and analytics decision makers.
Based on our definition of big data, what best describes your firm's current usage/plans to adopt big data technologies and solutions?

- Planning to implement within the next 12 months: 30%
- Implementing/implemented: 24%
- Interested but no immediate plans (within the next 12 months): 23%
- Expanding/upgrading implementation: 15%
- Not interested: 4%
- Don't know: 4%

Base: 3343 global data and analytics decision makers.
Four questions to ask yourself: are you ready (or not) for big data?

1. What are the typical analytics issues where big data is clearly not the answer?

2. What are the typical analytics issues and requirements where a different technology, not necessarily big data, may be the answer?

3. What are the typical business and technical requirements where big data may indeed be the answer?

4. What are the implications of upgrading analytics to big data?
Four questions to ask yourself: are you ready (or not) for big data?

1. What are the typical analytics issues where big data is clearly not the answer?

2. What are the typical analytics issues and requirements where a different technology, not necessarily big data, may be the answer?

3. What are the typical business and technical requirements where big data may indeed be the answer?

4. What are the implications of upgrading analytics to big data?
1) Business challenges you will NOT solve with big data

- Different points of view on who owns analytics.
- Disconnect on analytics goals and priorities.
- Treating data governance as a technology, not as people and processes.
- Data quality.
Business vs IT disconnect

Technology

› Single analytics platform

› Streamlined data architecture

› Centralized support

› Single version of the truth

Business

› I just want to get my job done

› Single version of the truth is not my top priority

› Good enough but timely data/info is good enough for me
Four questions to ask yourself: are you ready (or not) for big data?

1. What are the typical analytics issues where big data is clearly not the answer?
2. What are the typical analytics issues and requirements where a different technology, not necessarily big data, may be the answer?
3. What are the typical business and technical requirements where big data may indeed be the answer?
4. What are the implications of upgrading analytics to big data?
2) Non Hadoop technologies that can help you address business challenges

- Agile analytics platforms to improve self-sufficiency of business users.
- Columnar SQL databases to support frequently changing requirements.
- NoSQL databases to support complex data requirements.
- Analytics on analytics to improve business and technology alignment.
Over half of enterprise analytics content resides in home-grown shadow IT analytics applications.

% Responding

- 63% for 50%+
- 35% for 60%+
- 26% for 70%+
- 15% for 80%+
- 6% for 90%+
- 3% for 100%

% Homegrown Enterprise analytics Apps

Source: How would you describe the Enterprise analytics applications or tools that you use? Are they company-issued or “homegrown”? Base: 249 North American business decision-makers. Forrester’s “It’s Time For A User-Driven Enterprise analytics Strategy” report.
We Have Entered The Age Of The Customer

Winning In The Age Of The Customer

1900: Mass manufacturing makes industrial powerhouses successful.

1960: Global connections and transportation systems make distribution key.

1990: Connected PCs and supply chains mean those that control information flow dominate.

2010: Empowered buyers demand a new level of customer obsession.

Source: Forrester's “Winning In The Age Of The Customer” report.
10 dimensions of business agility

Source: Forrester’s “The 10 Dimensions Of Business Agility” report
Source: Forrester’s “The 10 Dimensions Of Business Agility” report
Source: Forrester’s “It’s Time For A User-Driven Enterprise analytics Strategy” report.
Four questions to ask yourself: are you ready (or not) for big data?

1. What are the typical analytics issues where big data is clearly not the answer?
2. What are the typical analytics issues and requirements where a different technology, not necessarily big data, may be the answer?
3. What are the typical business and technical requirements where big data may indeed be the answer?
4. What are the implications of upgrading analytics to big data?
3) When Hadoop is the right answer to address your business challenges

- Open up more data sources for your analytics applications.
- Justify Hadoop investments by reducing budgets allocated to proprietary systems.
- Turn your Hadoop data hub into a sandbox for business analysis and data scientists.
- Extend Agile analytics to big data with Hadoop on-demand data marts.
We use a small fraction of enterprise data...

But anecdotal evidence and deeper dives often show that less than 10% of unstructured and less than 20% of structured data are only being turned into information.

Base: 2094 global technology decision-makers who know how much analytics data their firm uses.
Note: These percents are estimates based on ranges reported. The values are not exact.
Source: Forrester’s Business Technographics® Global Data And Analytics Survey, 2016
Big data = scalability and agility
Do you know what schema on write vs schema on read is?
SQL

Data and metadata are tightly bound
= Schema On Write or Early Binding
Data and metadata are tightly bound
= Schema On Write or Early Binding

Data and metadata are separate
= Schema On Read or Late Binding

SQL

NoSQL or SQL on demand
Data and metadata are tightly bound
= Schema On Write or Early Binding

Data and metadata are separate
= Schema On Read or Late Binding

The European version 😊
Four questions to ask yourself: are you ready (or not) for big data?

1. What are the typical analytics issues where big data is clearly not the answer?
2. What are the typical analytics issues and requirements where a different technology, not necessarily big data, may be the answer?
3. What are the typical business and technical requirements where big data may indeed be the answer?
4. What are the implications of upgrading analytics to big data?
Are you REALLY sure you are ready for big data?

In the world of big data single version of the truth becomes relative and contextual.

Data in Hadoop data lakes will never be 100% clean and integrated.

Data governance must adapt to different stages in the life cycle of big data.
Sample hub-and-spoke BI architecture

Ingest

Move

Data management and integration

Consume

Internal data sources

External data sources

Data lake/data hub (distributed)

Local lakes

Data discovery accelerators

Hub

Spoke

Ad hoc interfaces

Standard interfaces

Virtual DW

Data exploration

BI

Metdata for date lineage/impact analysis

DW/DM-out-of-the-box

Data governance, BI on BI

“Cold” area: slow, less expensive Hadoop or Hadoop-like platform

“Warm” area: faster, more expensive DBMS

“Hot” area: fastest, most expensive in-memory

Source: Forrester’s “Boost Your Business Insights By Converging Big Data And BI” report
More Systems Of Insight best practices

- Contextual insights (embedded, pervasive analytics)
- Suggestive analytics
- Cognitive computing
- Insights to execution / action (actionable analytics)
- Continuous feedback loop for learning and improvement
Thank you

Boris Evelson

bevelson@forrester.com
http://www.forrester.com/Boris-Evelson
http://blogs.forrester.com/boris_evelson
https://twitter.com/bevelson
https://www.linkedin.com/in/bevelson
https://www.facebook.com/ForresterBI