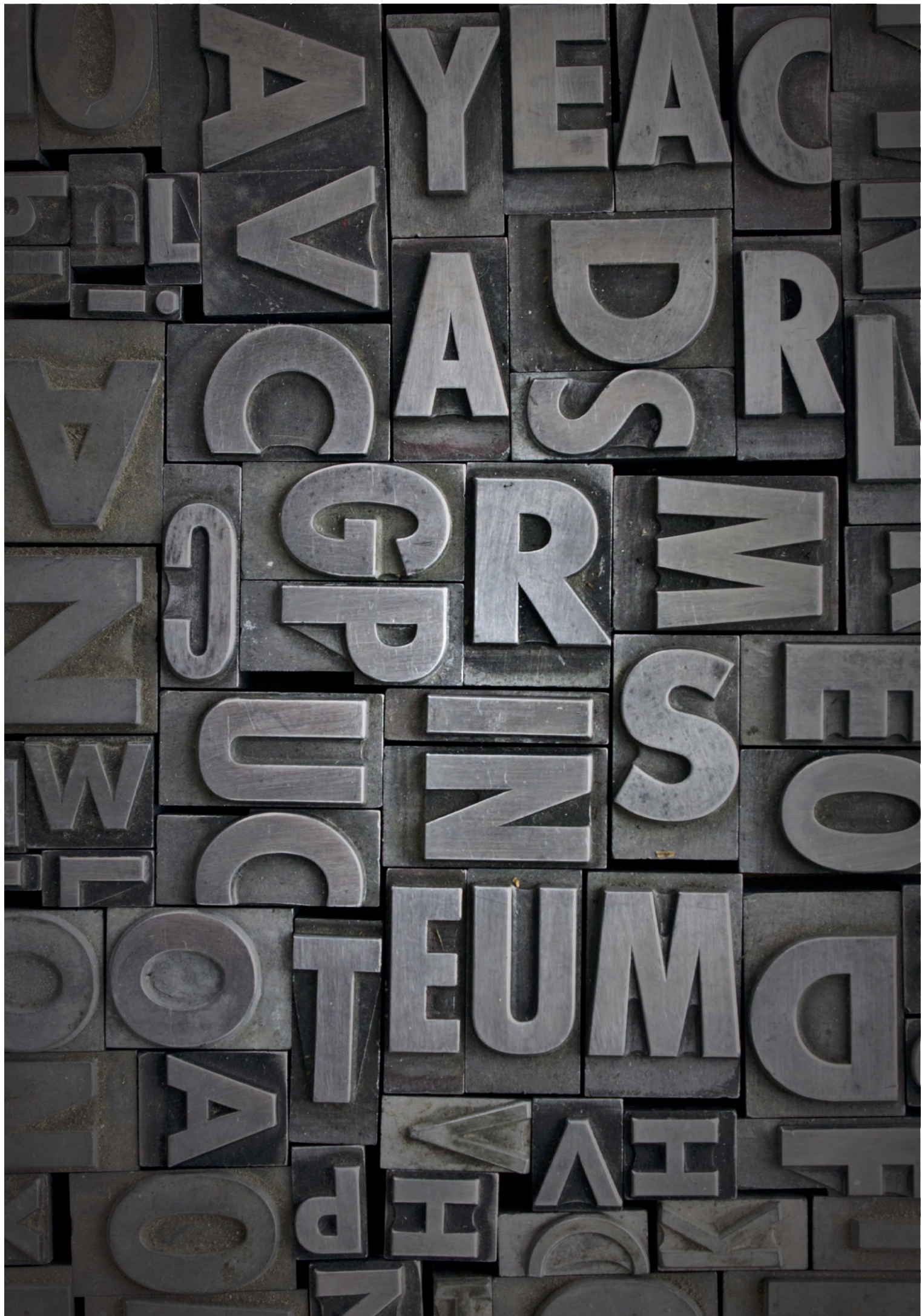




Ipsos Loyalty

# **A GUIDE TO TEXT ANALYTICS** **ALL YOU NEED TO KNOW FOR** **SUCCESSFUL DEPLOYMENT**

Fiona Moss | Jean-François Damais



# CONTENTS

---

04	INTRODUCTION
05	TEXT ANALYTICS:WHAT'S THE BIG DEAL?
07	THE IPSOS APPROACH
12	CONCRETE TEXT ANALYTICS APPLICATIONS
13	CASE STUDIES
14	THE FUTURE?



## INTRODUCTION

Text analytics uses a variety of computer processing techniques to identify and quantify the main themes and sentiment in text data. There has been a significant increase over the last few years in the volume and variety of sources of unstructured data, including feedback from customers, potential customers, employees, members of the public and information systems. The huge value that lies buried in this data means that the ability to extract actionable insights and intelligence from unstructured text is now crucial for most organisations.

Whereas the potential to derive insights from this increasingly performant technology is huge, there are a number of pitfalls organisations need to avoid. One of the most dangerous is the belief that technology in itself, regardless how state of the art, is enough to derive good and actionable insights.

Analytical expertise is needed to set up and carry out the analysis in the right way, but also to interpret, validate, contextualise and link text analytics results to other sources of data in order to disseminate the right insights to the right people within organisations.

For any organisation considering embarking on a text analytics exercise, this requires a conscious leap of trust. We therefore argue against killing the analyst just yet as without the expertise and investigative mind of an expert results are likely to fall short of expectations.

# TEXT ANALYTICS - WHAT'S THE BIG DEAL?

Text analytics is definitely a hot topic these days and rightly so. When Ipsos started to provide text analytics services to its clients in 2009 this capability was very much a niche offering and seen by most organisations as an added value and nice to have. Demand is now at an all-time high and growing across most industry sectors: at Ipsos the amount of text analytics work has grown by 70% over the last year following a 70% growth the year before.

## There are significant benefits to text analytics

- **Cost:** huge economies of scale can be achieved versus manual (human) coding, particularly when it comes to large and recurrent datasets that need to be analysed. Once the text analytics framework is set up and validated, the process is pretty much automated – although a degree of human intervention remains necessary to ensure the ongoing accuracy and relevance of the text analytics framework.
- **Speed:** by the same token, huge efficiencies can be achieved versus manual coding post set-up.
- **Consistency:** the analysis is more consistent than manual coding across waves or across targets, meaning changes or variation in the results represent variation in what is being said, rather than being the result of a different coding approach.
- **Scalability:** once a framework is set up for a particular vertical or business issue, data from other waves or different projects can be analysed very efficiently using the same framework.
- **Shorter questionnaires and better respondent engagement:** a single, well-analysed open-ended question can deliver more insight in one go than asking respondents to rate a whole battery of service or product attributes for example. This can mean cutting the number of questions asked, thus reducing the risk of respondent fatigue.
- **Data integration:** text analytics has a practical Big Data application. Different data sources can be analysed using the same framework increasing our ability to compare topics across various data sources and make relevant linkages for better predictive analytics.
- **Better insights:** text analytics is not just a cheaper alternative to manual coding. It also significantly increases our ability to extract meaning, intelligence and predictive/explanatory power from text data on a large scale. Examples are given later in this paper.

**Text analytics demand has grown by 70% over the last year**

## There are pitfalls too

As with any analysis, there are drawbacks and considerations. First and foremost, it is important to be realistic – text analytics is not a perfect solution, and even with the best analysts, there is an inevitability that a small proportion of comments will be incorrectly categorised.

For businesses willing to look past this, the benefits of text analytics are manifest. Beware, however, the analyst who guarantees 100% accuracy – the goal here is 'good enough', and the promise of perfection suggests that the analyst does not recognise the potential difficulties inherent in this analysis. Better trust the analyst who points out where a problem may lie and suggests a way forward, than the analyst who says there are no problems.

## The most common pitfalls are:

- **Beware of the hype:** buzz words – machine learning, NLP, concepts, target terms, queries, Big Data, to name but a few – are everywhere, but what do they mean? Getting past these and into the detail of what the tool actually does, and how the analyst interprets this, is essential.
- **Beware unrealistic expectations:** there is no miracle "press a button" type solution, and being realistic about the time investment from your analyst, and your own time investment to think about and act on the results is key. During discussions with text analytics suppliers all of them claimed to have the best technology available in the market – something that evidently cannot be the case. Clearly though, this can be confusing for any organisation trying to implement a solution.
- **Beware of information overkill:** text analytics is a means to an end. The challenge is to turn large volumes of heterogeneous data into knowledge, intelligence and insight – and this is not easy. Numbers without insights can be a common problem in text analytics.
- **Managing multiple languages:** different tools have different capabilities for dealing with multiple languages. Knowing in advance that your research is likely to involve multiple languages will help inform your decision about which tool to use or approach to adopt: these may range from setting up a text analytics model in each language, to translating all of your sources into a single language for use with a tool that has only one language capability.

Demand for text analytics is at an all-time high and the number of applications is growing fast

Data Type	Examples	Applications
Verbatims from quantitative surveys	<p>Brand trackers, customer/employee engagement studies</p> <p>Advert/Product testing likes &amp; dislikes</p> <p>Enterprise Feedback Management/Voice of the customer programmes</p>	<p><i>What are people saying?</i> Auto-coding of verbatim into themes</p> <p><i>What is good and bad?</i> Measure and track sentiment overall and by themes</p> <p><i>What should we improve first?</i> Text based impact/driver analysis for action prioritisation</p> <p><i>Real-time analysis of customer feedback</i> Prioritise actions and interventions in timely fashion</p>
Qualitative surveys	<p>Online communities, focus group transcripts, public consultation transcripts</p>	<p><i>What is being said and in what context?</i> Extract key words and patterns in data</p> <p><i>Can we map the key themes?</i> Visualise patterns and relationships between concepts</p> <p><i>What should wider analysis focus on?</i> Guide/focus analysis</p>
Social media data	<p>User Generated Content from social networks, data from forums/blogs/review sites, news sites etc...</p>	<p><i>What are people saying?</i> Auto-coding of comments into themes</p> <p><i>What is good and bad?</i> Measure and track market sentiment</p> <p><i>What is being said about our competitors?</i> Competitive intelligence</p> <p><i>Is there anything new we should be worrying about?</i> Identify early signals of user/customer opinion</p>
Feedback/data held by organisations	<p>Call centre logs/recordings, data from website, email data etc..</p>	<p><i>What have people said in the past?</i> Audit of existing data</p> <p><i>Is any of our existing data relevant for our current task?</i> Find relevant data for further analysis</p> <p><i>What are people saying?</i> Identify recurrent themes</p>

Moreover, many tools can merge sources and/or deliver on a number of these applications simultaneously. The key is knowing this in advance so that all set up can be appropriately put in place. This means that text analytics can be a perfect solution for data integration or Big Data applications.

# THE IPSOS APPROACH

## Using the right tools in the right way

Trying to set up a text analytics process can be quite daunting. The market is very fragmented with many different players offering text analytics services – from research organisations offering full service (analysis and interpretation) to pure players, social media analytics boutiques or open source software and university spin offs.

Since 2009, Ipsos has been conducting ongoing research and development into text analytics. This started as a

simple validation of text analytics as a viable, reliable and cost effective alternative to manual coding.

As our knowledge of text analytics has grown, along with the sophistication of the analysis techniques undertaken, we became increasingly aware of the need to evaluate the different tools in the market.

During 2014 we conducted a comprehensive review of many of the tools currently available and identified pros and cons for each.

Approach	Pros	Cons
<b>Semantic/linguistic</b> Tools that leverage semantic networks, dictionaries, linguistics rules and rules-based categorisation techniques	<ol style="list-style-type: none"><li>1. Extracts meaning and sentiment</li><li>2. Granular classification</li><li>3. Scalable (taxonomies, dictionaries)</li><li>4. Transparency and control</li></ol>	<ol style="list-style-type: none"><li>1. Time consuming to set up</li><li>2. Nuancing, sarcasm difficult</li><li>3. Analytical and language skills needed</li></ol>
<b>Supervised machine learning/human training</b> Tools that require human input/ coding to train the system and create an automated model that can then code similar type of data	<ol style="list-style-type: none"><li>1. Quick for simple classification</li><li>2. Easy to set up</li><li>3. Can work well to capture abstract dimensions</li></ol>	<ol style="list-style-type: none"><li>1. Lacks granularity</li><li>2. Black box</li><li>3. Not very scalable</li><li>4. Human bias</li></ol>
<b>Unsupervised machine learning/statistical</b> Tools using probabilistic models to group words or phrases based on co-occurrence and proximity patterns	<ol style="list-style-type: none"><li>1. Instant</li><li>2. Great for exploration of large heterogeneous datasets</li><li>3. Great data viz</li></ol>	<ol style="list-style-type: none"><li>1. No control</li><li>2. Hard to apply conceptual framework</li><li>3. Results not always meaningful ("data driven")</li></ol>

Although each of the tools presented us with different strengths, challenges and functionalities, we gained the following learnings:

- There is no perfect technology. Knowing the strengths and weaknesses of the technology used is key to getting valuable results.
- There is no miracle “press a button” type solution, even the best tools need some human intervention and analytical expertise; press button solutions rarely yield valuable insights. Expert intervention turns simple text mining into meaningful analytics that can be used to inform business strategy.
- There is no “one size fits all” tool – depending on the type of data or requirements some tools and technologies might be better suited than others.

Ipsos Loyalty's main approach leverages powerful semantic processing and parsing of language together with advanced statistical analysis. This approach enables us to extract meaning, sentiment and explore the relationships between topics in the data – “what the data is saying”, but also to robustly quantify the occurrence of topics and the strength of the relationships between key topics and any other type of structured data including survey KPIs.

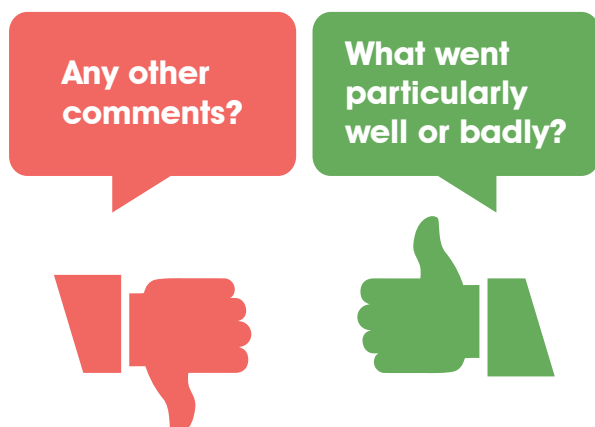
Ipsos Loyalty has a portfolio of text analytics tools that enables us to use the right tools in the right way for any given situation.

### Asking the right questions – and asking them in a way that plays to text analytics' strengths

One of the essential steps to creating actionable results is having a clear idea of the question that you would like text analytics to answer. This is particularly the case when it comes to survey data, where the right kind of question can make all the difference in terms of the quality of the results.

Taking an example of a survey about customer service, asking ‘any other comments’ will generate a broad range of comments that lack detail and actionability. However, changing the question to ‘tell us how to improve customer service’ will orientate the comments, allowing text analytics to generate a more focused model with more structured categories and subcategories and, hence, more insight.

#### Asking the right questions



In social media, identifying the question is no less important – but here it is more about deciding what you would like to learn from the sources available. For example, needing to know what is being said about a particular product on a particular forum would result in a very different text analytics structure than that required for analysis of customer service comments relating to a given retailer for example.

Simply conducting text analytics without having any explicit questions can provide disappointing results either because they are too diluted (e.g. trying to cover too many topics or too broad a field) or produce results that are not actionable or not of interest.

Without good quality, focused verbatim data, and/or a clearly identified question to answer, the results may leave the user feeling they have a fairly blunt tool that may give a sense of what is being said, but does not deliver any real insight or actionable findings.

### Customizing the text analytics approach

Alongside asking the right questions, it is important to bear in mind the set-up of the text analytics tools. Many providers offer a generic framework for creating a categorisation; but this, in much the same way as asking a vague question, can result in unfocused findings. For example, categories that reflect 90% of the comments analysed and consequently do not provide enough granularity to deliver any real understanding of what is said; or conversely a high volume of categories that contain only 1-2% of the comments, and as such are scarcely more digestible than reading the full range of comments without text analytics.

**Ipsos Loyalty uses a customized approach, basing the categorisation structure/framework around the contents of the comments and the question to be addressed**

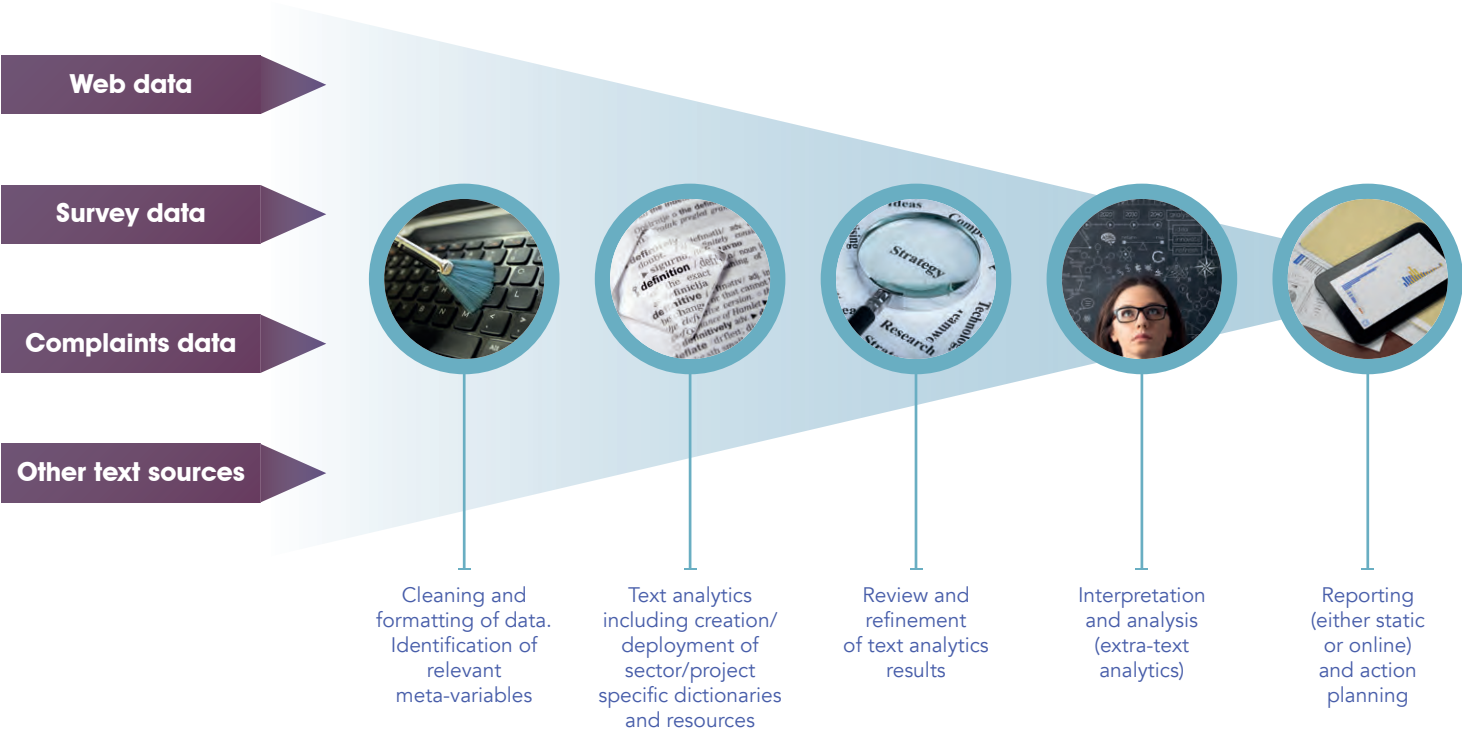
To overcome this, Ipsos Loyalty uses a customized approach, basing the categorisation structure/framework around the contents of the comments and the question to be addressed.

This ensures a meaningful categorisation when looking at the results, such that they deliver insights and are manageable in terms of the volume of data generated.

Putting the analyst at the heart of it all:  
Choosing the optimum combination of automated - and analyst - run techniques

Text analytics involves multiple steps to both identify the content of the text data being analysed and then derive additional insight from it.

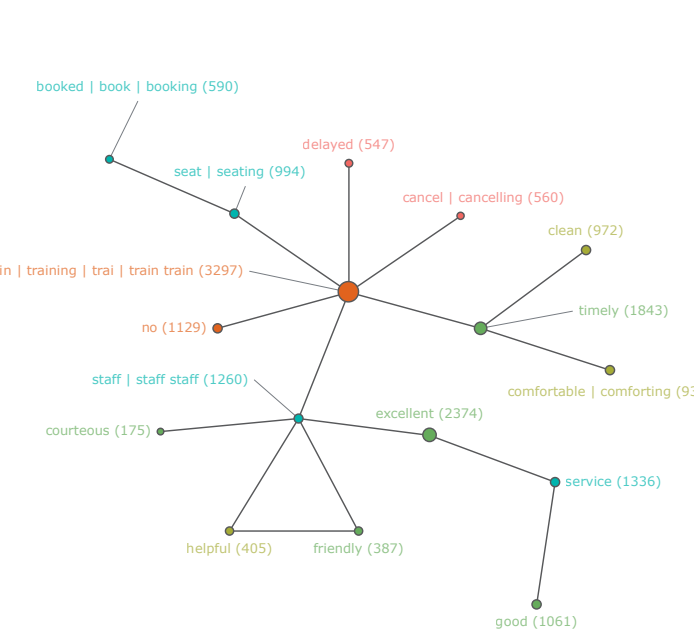
The text analytics process - the analyst is essential at each step



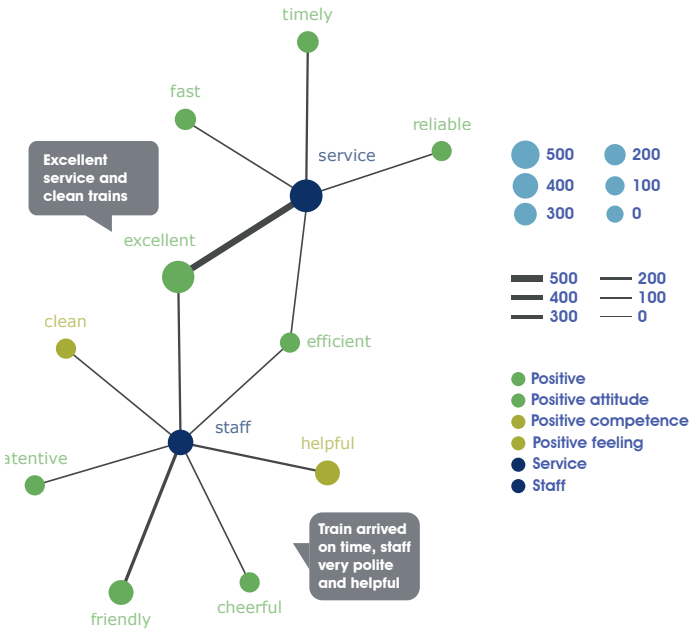
Exploratory analysis

Text analytics starts with exploration. In the exploration phase key relationships between the topics in the data are identified and represented graphically using maps. This analysis can be used to inform the broader categorisation of the data into themes and provides useful context by allowing the user to explore what has been said alongside those themes. It helps to create a wider picture of what the customer has been saying, and so build up a greater depth of understanding of the customer experience.

Exploration webs

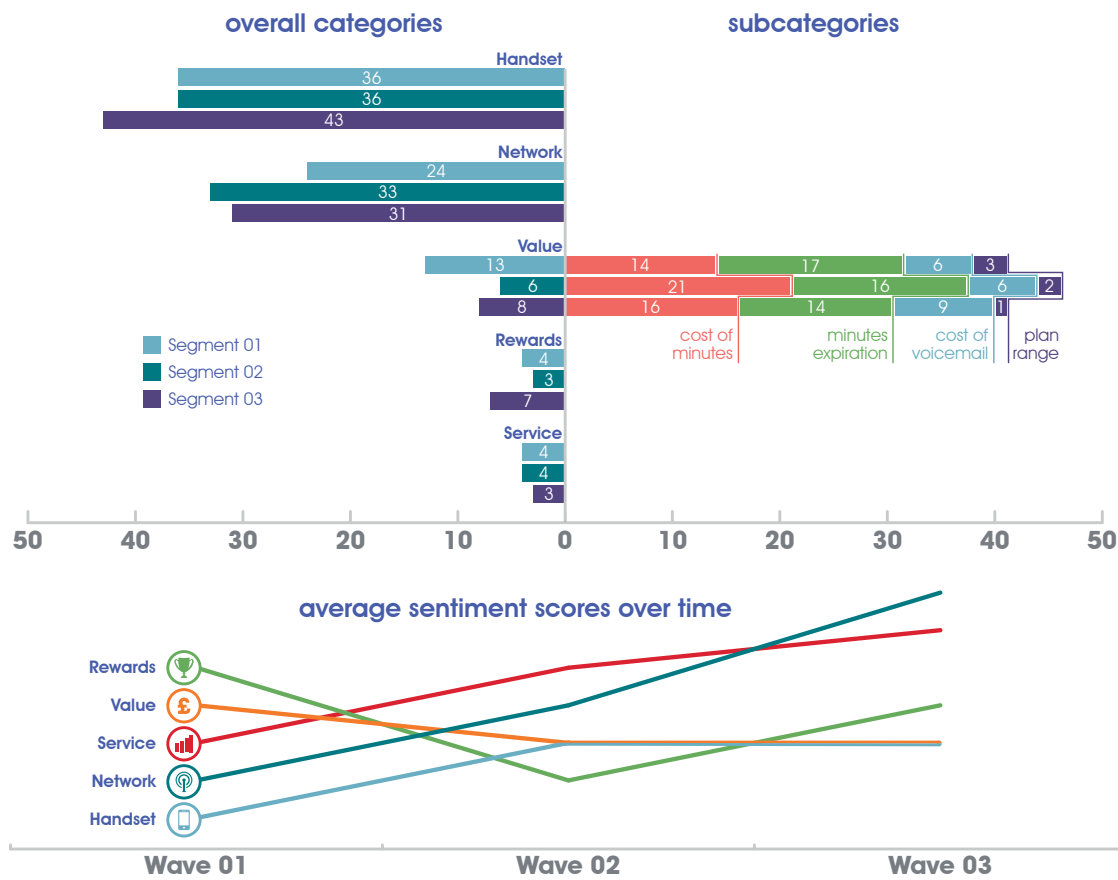


Example of positive patterns in the data  
Positive comments - service and staff

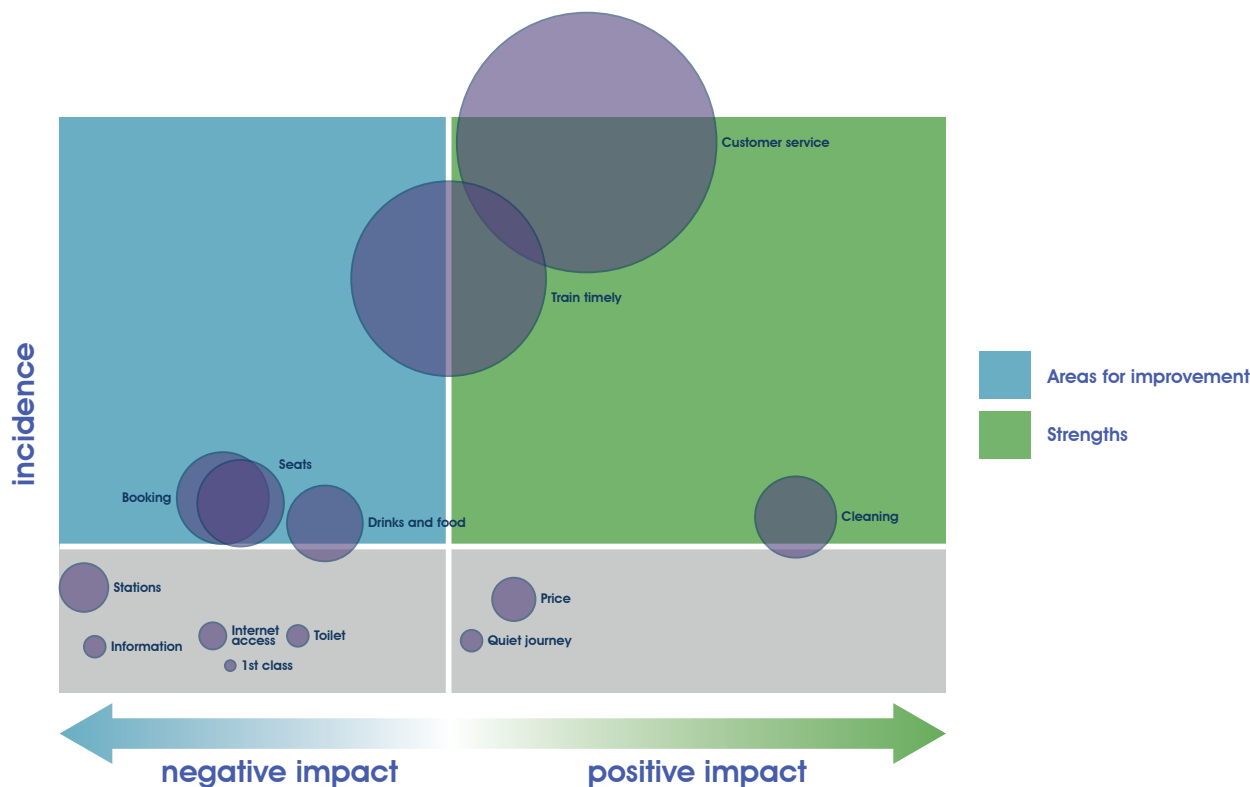


Quantification

The analysis then moves on to quantification. Hierarchical classification of unstructured data into themes identifies the most common pieces of feedback given – and whether this is praise or criticism (sentiment can be measured). Results can be split by different customer groups and tracked robustly over time.



In addition we can produce text based impact analysis to quantify the relationships between themes and KPIs (e.g.NPS). This can build a picture of which categories pull down the KPI score the most, and which drive it up. By factoring in the volume of people who have mentioned a particular theme/category we are able to identify priority improvement areas (i.e. those that need urgent attention) as they will have the biggest negative impact on KPIs and/or extend over the widest range of customers.

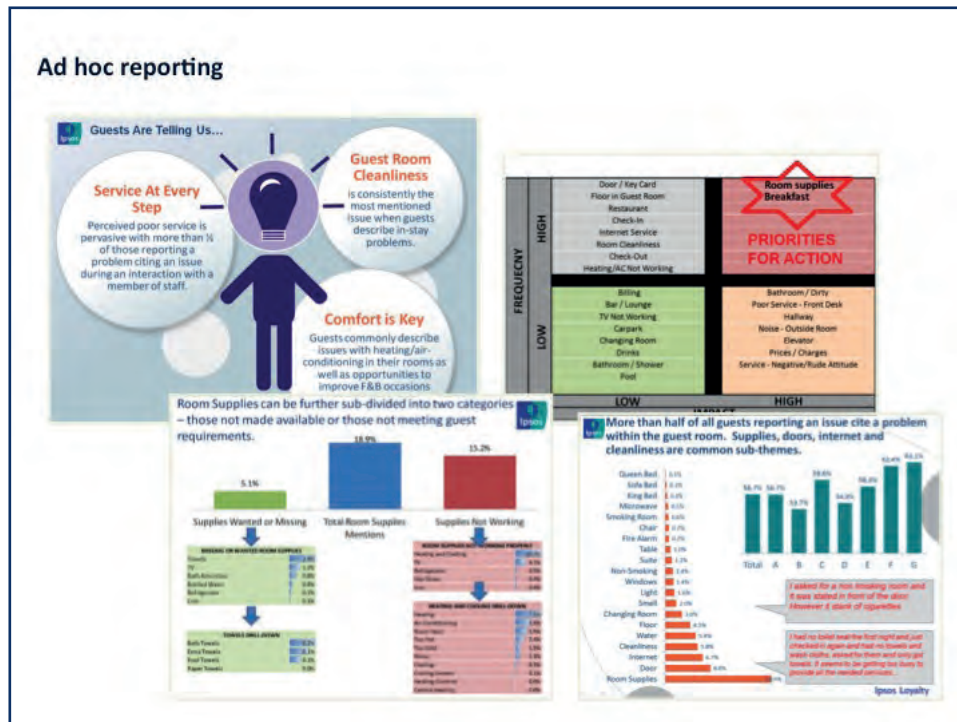


## Dissemination

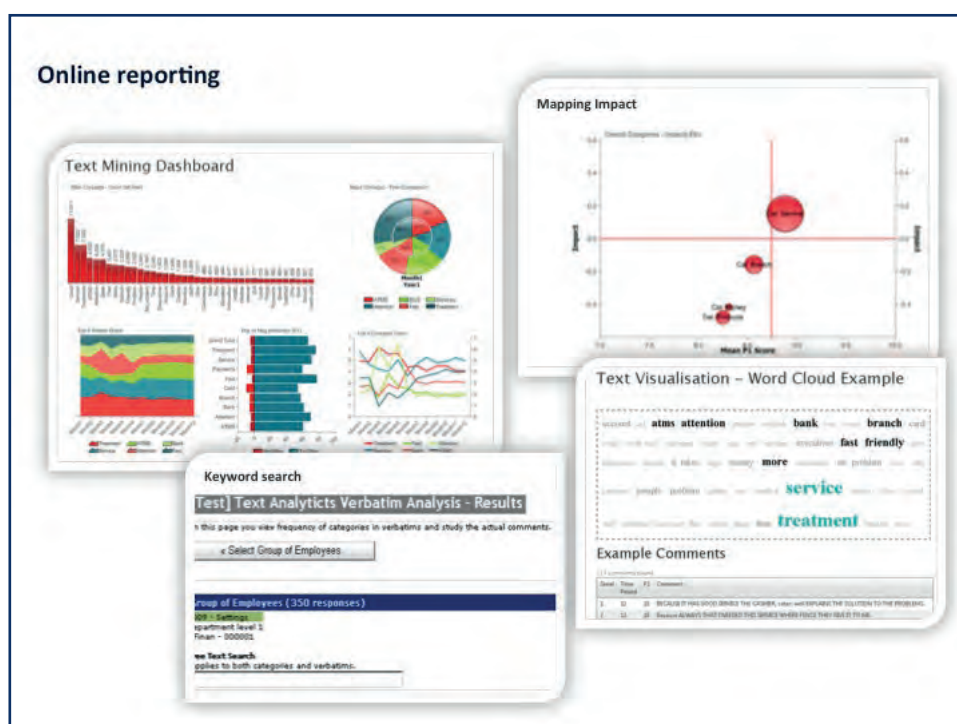
The final stage of text analytics is dissemination – key to get right for action to be taken.

There are two main types of reporting:

Ad-hoc Voice of the Customer insight reports which summarise the key findings of the text analytics and bring the voice of the customer to life. These reports can be produced every quarter alongside traditional reporting or can be produced to focus on particular topical issues as and when required.



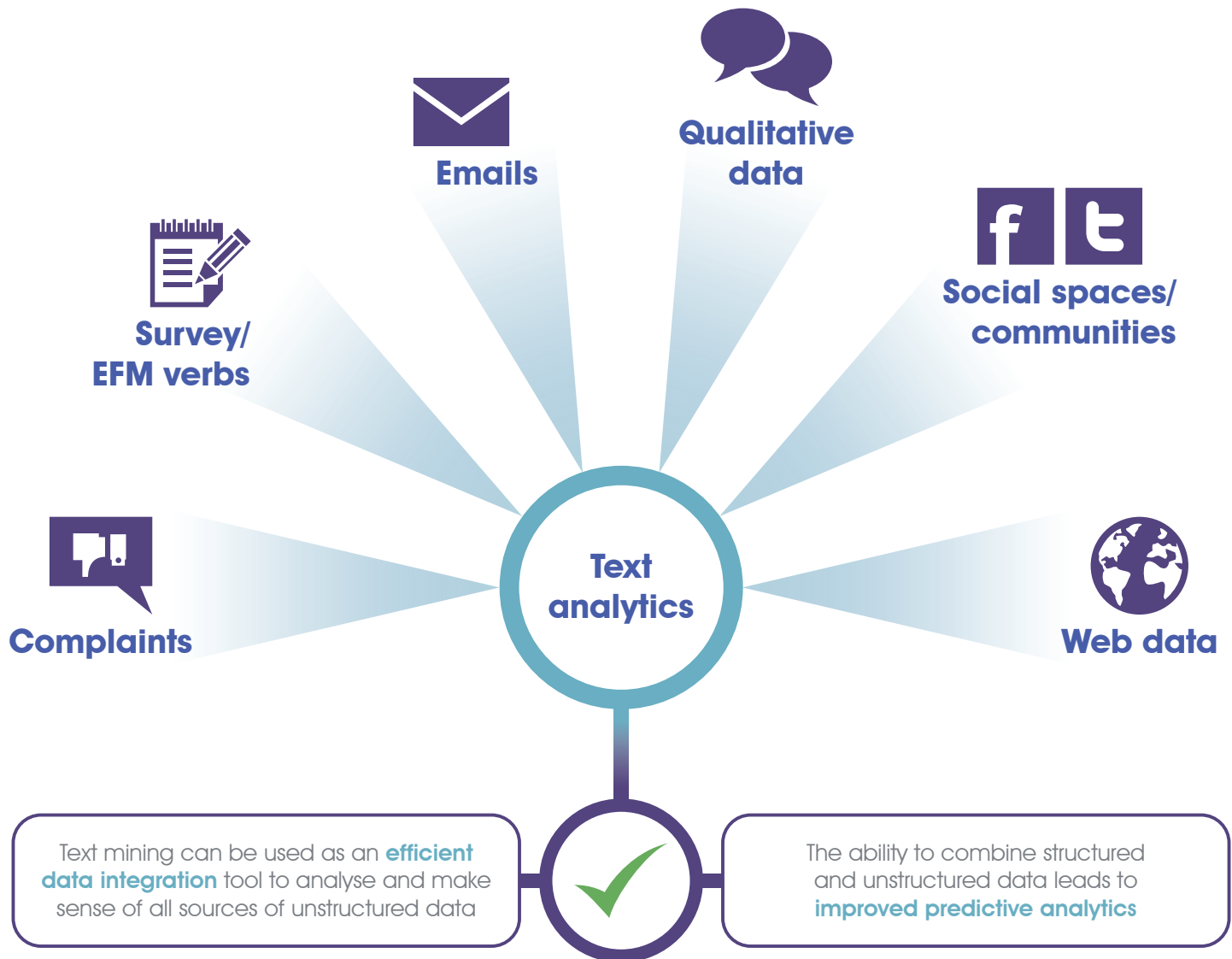
Online reporting of 'voice of the customer' text analytics outputs to various stakeholders within an organisation. Results can be uploaded periodically or in real-time depending on requirements.



# CONCRETE TEXT ANALYTICS APPLICATIONS

## Data integration

Today we receive verbatim or written information from a multitude of different sources. But what happens when we want to



## We have built up experience working with a number of different data types

- 1** Real-time analysis of customer comments as part of an EFM/transactional research programme to provide ongoing/tracking findings over time, quickly and cost effectively. This approach is typically process driven with a goal to identifying quick interventions to improve the customer experience on the back of transaction feedback or EFM results.
- 2** In-depth detailed analysis – particularly where large banks of text data are held – in order to better understand the customer experience and identify strategic actions to improve the relationship with them.
- 3** To combine data sources such as social media and survey data to build a more holistic picture of the customer point of view.

# CASE STUDIES

1

## Manufacturer



Over the course of two years this project involved analysing verbatim comments from customer satisfaction research from 31 countries and in 26 languages, leading to the ongoing analysis of several hundred thousand comments a year. We text mined the verbatim comments and presented the results in an interactive portal, allowing the client to interrogate their own data. The primary objective was to help the client to understand what their customers were saying and identify the recurring themes. To do this, the analysis had a high level of granularity – so that the client could identify the main causes for customer comments (eg customer service, after sales care) but also the more detailed elements (eg specific faults). This enabled the client to target improvements in its offering to specific areas.

2

## Automotive dealer



We worked with an automotive dealer to better understand customer comments left in their satisfaction study. Given the volume of comments available, we used text analytics in order to benefit from its economies of scale. The outputs from the text analytics also lent themselves to a number of further statistical analysis techniques to help deliver insight from the comments.

The analysis took place in two phases. For phase one we built a hierarchical categorisation of the comments to understand what was being said. This included large main themes organised around the business's structure and numerous more detailed categories to deliver actionable findings for specific teams in the business. We then ran impact analysis – or text based drivers analysis – to identify the priority improvement areas for our client.

For phase two, we deep dived into customer service using a sentiment engine and a correlation analysis to build a picture of good and bad service, before finally assessing our client's performance against this. The findings from this fed into business and training planning.

3

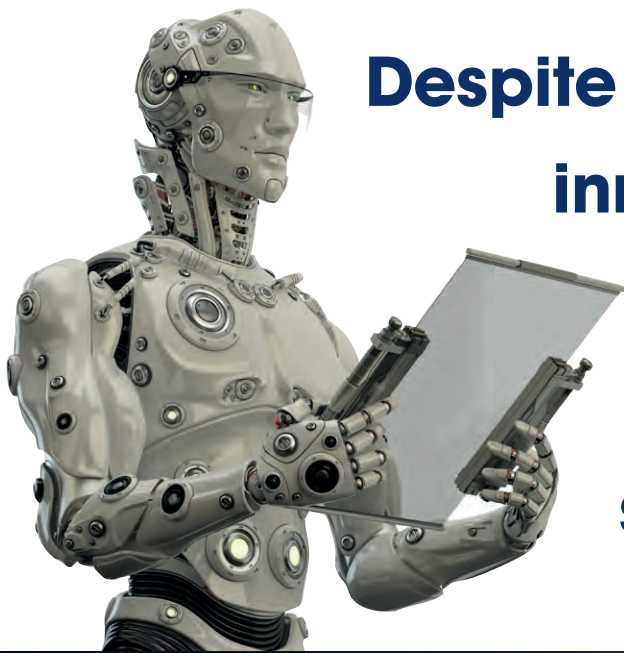
## Mining the social media morass



We used text analytics to mine seventy thousand Twitter comments. We established a correlation between Share of Voice and Market Share and identified pull factors (what attracts customers) and push factors (what detracts customers) across the competitive set.

We were then able to track mentions over time and built longitudinal stories to add context and depth to more traditional survey research insights.

# THE FUTURE?

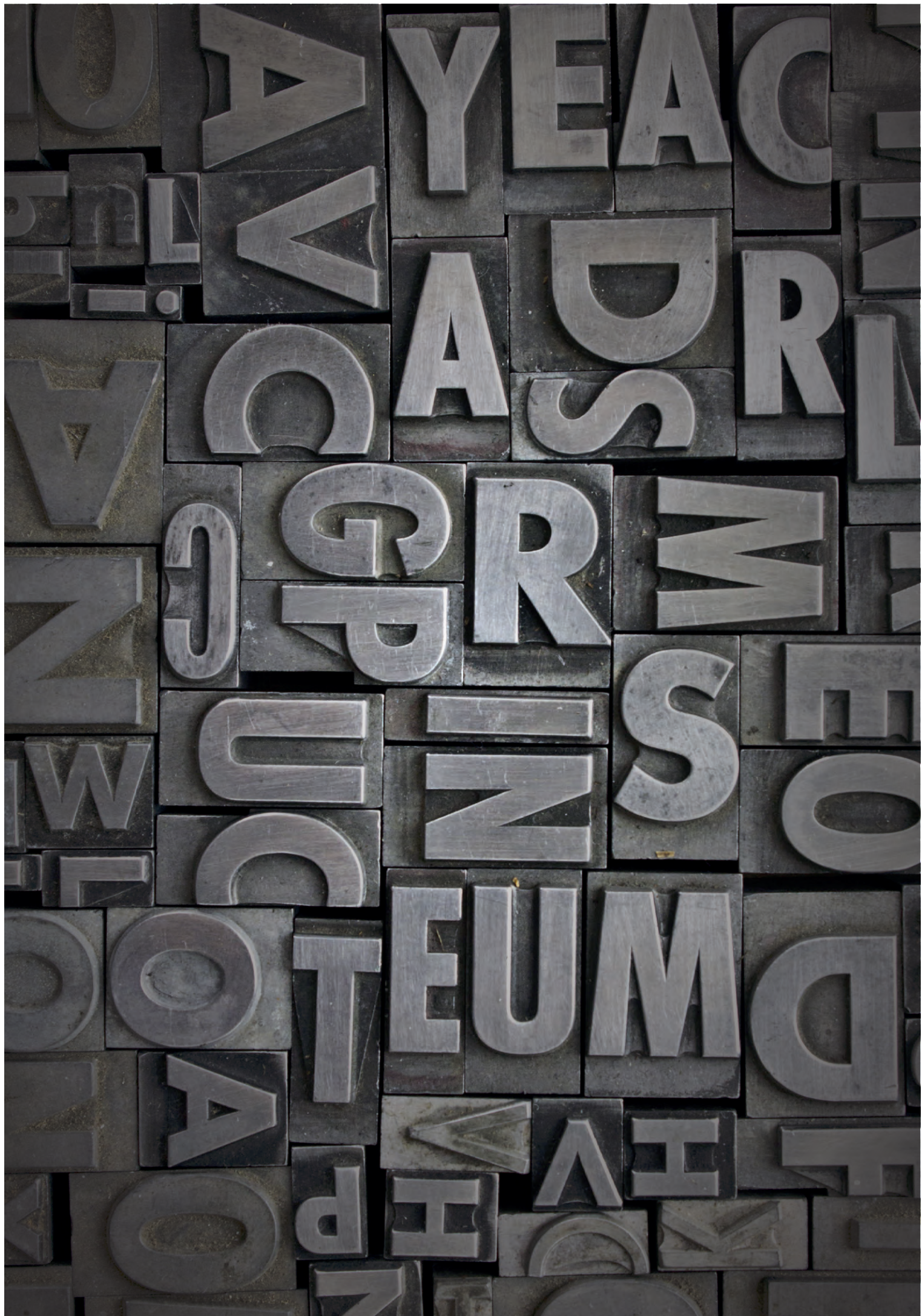


**Despite constant technological innovation in text analytics, we believe the analyst will remain essential to successful deployment.**

Text analytics providers are stepping up to this challenge providing tools with increasing sophistication, language capabilities and accessibility. Along with these improvements are increasing promises and hype about what can be delivered and the level of insight available. However, there are also more niche tools, more jargon and a more competitive terrain where all providers are jostling to be best in class. Navigating this terrain to get to a sensible offer that delivers results has therefore become increasingly risky for businesses starting out on their text analytics journey.

It is, however, a terrain that we have explored and it is our belief that – despite the temptations and promises of push button/fully automated solutions - truly actionable results will continue to be delivered by text analytics tools driven by trained analysts. It is only through this approach that the strengths of each tool can be fully exploited, the framework oriented towards business issues, the results validated and consultancy given on interpretation and action planning. And until text analytics software succeeds in replicating the human brain, this means keeping the analyst very much alive and at the centre of text analytics.

Technology is important, but it is a means to an end. It is the knowledge of the data, how to manipulate and interpret the results and how to tailor these to the individual business questions that leads to truly actionable results. This places the analyst at the heart of the analytics.



**Ipsos MORI**

3 Thomas More Square  
London  
E1W 1YW

t: +44 (0)20 3059 5000

[www.ipsos-mori.com](http://www.ipsos-mori.com)

@IpsosMORI