

# Cold, hard numbers and bright insights from customer comments:

## Text analytics for marketing and customer experience.

### Why text analytics? Why now?

When customers are freed to express their thoughts in their own words, they reveal more about their motivations: they give marketing and customer insight teams the “Why” to structured data’s “What?” In the last 3 years, text analytics has successfully made the leap from university supercomputer to the office desktop and it’s now possible to analyse 10,000 or a million customers’ comments with full statistical reliability in near real-time, and while maintaining enough granularity to provide qualitative insights.

In this paper, we give an overview of the market today, we share knowledge from a selection of consultants’ work with multi-national brands over the last 3 years, and we offer advice on how to avoid expensive dead-ends.

Finally, on the back page you’ll find a crib sheet of those text analytics terms which we dearly wish someone had given us, when we started this journey.

### Introduction

Text analytics has been around for a long time, but only recently has it become automated. In the diagram below, we show the broad categories of different methods. Manual coding is still taking place, of customer satisfaction surveys for example, but it is time consuming and its accuracy depends on the day-to-day operations of the team. Automated methods are quicker, more reliable and are now widely preferred.

Text analytics can analyse customer comments from any source, including CSat surveys, social media, email and contact centre transcriptions. A recent IBM survey found that predictive analytics now feature most highly on CMOs wish list, with more than half already involved in using analytics to capture customer insight. <sup>1</sup>

The number of text analytics providers has grown rapidly and software is now available that can be managed in-department, on a range of budgets, with varying levels of disruption, and in time frames that range from ‘right now’ to ‘still waiting’.

## Different text analytics methodologies

### MANUAL

#### HAND CODING

- +: accuracy
- : time-consuming
- : limited volumes
- ?: variable quality

#### STATISTICAL

“word counting”

- +: simple
- : simplistic\*

### AUTOMATED

#### LINGUISTIC

“a grammar machine”

- +: higher accuracy
- +: fast

#### MACHINE-LEARNING

“give it a training dataset ... and time”

- +: insight
- : time-consuming
- : can’t spot outliers

\*man bites dog = dog bites man

<sup>1</sup>: “CMO insights from the Global C-suite Study” IBM Institute for Business Value, [pub 2014]



## Four ways text analytics is now adding value

Text analytics is now being used by customer experience and marketing teams to add value in four distinct ways:

1. **Reducing churn** by identifying early warning signals in customers' language;
2. **Answering the questions** which existing structured data has already raised;
3. **Being first to spot the added-value opportunities**, by hearing what customers are starting to talk about, before the conversation reaches critical mass;
4. **"Bright Lite Insights"** – text analytics is being used to drive single-issue, fast turn-around research; the resulting "quali-quant" findings are being used to prove the viability of bigger commercial projects.

## Insight goes HD

Unstructured data constitutes between 70% and 90% of all data available to brand teams. It's also the single fastest growing source. <sup>2</sup> Better than mere word counting (e.g. 'Word Clouds'), text analytics gives robust statistical insight into recurring *themes*.

A recent report by a leading industry analyst shows that marketing and customer experience directors are now using text analytics on a regular basis and the market is growing by 25% year-on-year. <sup>3</sup> Without it, the CEO is being given an incomplete picture.

## How text analytics has been used in the last year

A sample of case studies show the depth and breadth of value that text analytics is now adding:

(N.B. We have no commercial affiliation with any of the providers mentioned. Our role, after the recommendation of best-fit software, is the interpretation of the results of that software.)

- **Resolving confusing NPS:** In the States, Anderson Analytics helped Jiffy Lube resolve why sales didn't correlate with NPS scores. The team processed 400,000 survey responses, identified recurring themes in customers' comments and showed which of these most closely related to the sales and CRM data. <sup>4</sup>

- **Qual meets quant :** Clarabridge worked with a US airline to better capture the Voice of the Customer, analysing 7 million web-based surveys. The use of text analytics meant that open-ended questions could be used in the survey, which in turn meant that fewer questions needed to be asked; once the survey was shortened, more people responded. From the new data, the client discovered previously unsuspected themes of customer dissatisfaction and was able to dig deeper into their customers' true view of the brand.
- **Correcting customers' perceptions of staff:** At Verbal Identity, we worked with a UK 'big 4' supermarket to identify underlying causes of customer dissatisfaction <sup>5</sup>. We analysed 100,000 monthly customer comments and provided the customer experience team with, for the first time, a statistical weight to show how much "staff rudeness" was affecting NPS scores. Then, using linguistic analysis of individual responses, we discovered that it was the customers' inaccurate perception of the skill level of the staff which was driving those feelings about the quality of their interaction. Insights aren't insights unless they're actionable, so as a final step we created new staff titling and language programs that could radically alter customers' perceptions.<sup>6</sup>

## The 5 Basic Requirements for a text analytics project to be commercially usable:

In our experience, text analytics projects need to meet some real-life criteria before they can be considered commercially viable:

- **Repeatable:** Until your text analytics system has better-than-human levels of consistency, no one can trust it to give a dependable measure of changing company performance.
- **Timely:** If you can't analyse the current data before the next round of data arrives, either you're missing an opportunity or you are wasting money collecting too much data.
- **Robust:** Commercial text analytics software should be able to process all of your customers' comments, wherever they come from.
- **Zero or low disruption:** Your first text analytics projects should be run in-department and without requiring long build times: expect systems to be up and running within 8 weeks.

<sup>2</sup> [http://ikt.hia.no/perep/eip\\_ind.pdf](http://ikt.hia.no/perep/eip_ind.pdf)

<sup>3</sup> [http://breakthroughanalysis.com/2013/12/30/aw-re-aw-text-analytics-industry-study\\_start-ups-and-aquisition-activities\\_max-breitsprecher/](http://breakthroughanalysis.com/2013/12/30/aw-re-aw-text-analytics-industry-study_start-ups-and-aquisition-activities_max-breitsprecher/)

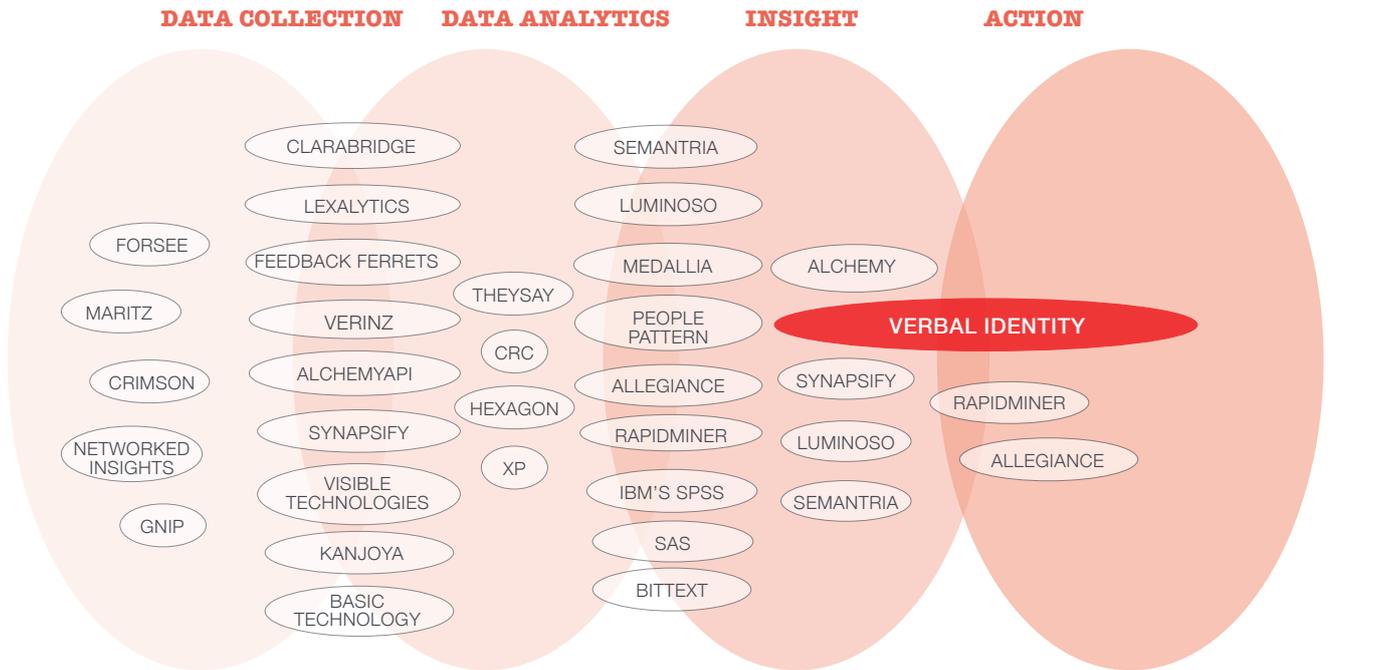
<sup>4</sup> <http://adage.com/article/dataworks/jiffy-lube-net-promoter-score-goose-sales/243046/>

<sup>5</sup> <http://semantapi.com/what-do-you-think-of-this-article/>

<sup>6</sup> We selected TheySay as software provider for this work



# An (incomplete) universe of text analytics software providers



Verbal Identity operates across this space, using linguistics to interpret results produced from other suppliers' software

## Expensive dead ends to avoid

**“Hire a monkey to flip a coin.”** A recent study shows that sentiment analysis is no better than chance: most platforms will accurately score only 50% of verbatims.<sup>7</sup> From our knowledge of how language works, we don't expect any improvement. Ever.

**“Health Vs Wealth.”** We see the relative values of ‘dashboarding’ Vs ‘actionable insights’. In our opinion, the constant monitoring of several metrics does give the health of the patient; but it's only once you focus on the causes of the symptoms that you can do anything to help them get better.

**“If it takes more than a month to build or 15 minutes to explain, something is wrong.”** Unfortunately, we are still meeting marketing or customer experience teams of 2 or 3 people who have spent many months building dictionaries of specialist terms for their expensive text analytics software. The best software can now ‘digest’ Wikipedia or other contemporary sources automatically, freeing teams to get on with analysis.

**“Excel never told you what to do, Word never wrote a great novel.”** The greatest advantage of text analytics isn't the data it provides: it's the opportunity to understand the meaning wrapped within the data. Nietzsche said “Every word is a prejudice.” And once you listen to *how* something is said as much as *what* is being said, you start to understand the real motivations of the speaker.

<sup>7</sup>: <http://www.verbalidentity.com/major-finding-outlook-negative-sentiment-analysis/>

## How to start with Text Analytics

We have been involved with a number of text analytics projects and discovered that the most successful outcomes depend on the following steps:

- 1. Start with a ‘conversation of possibilities’:** Work with people who understand both text analytics and your department's objectives. Spend an hour or a morning sharing knowledge and objectives, to identify together what this project *could* and *couldn't* discover.
- 2. Set clear goals:** You will always get interesting results from a text analytics project; sometimes, too many of them. Start with a focus on how any insights will have a clear impact on your company's top and bottom line.
- 3. What can you afford?** This isn't just a question of money, it's also about how disruptive you can afford your new text analytics processes to be.
- 4. Find the best fit software:** Pick the right software for your ends. This depends on your budget, your timings, your data and your desired outputs – not the software's.
- 5. Data is nothing without interpretation:** Look at how you can interpret the results. Admittedly biased, we see great linguistic interpretation helping to make insights actionable.
- 6. Jump, don't dive:** Conduct a brief Proof of Concept project to see how quickly and efficiently text analytics can give you actionable insights, and use this to generate consensus around the most effective route of further work.



# Jargon Safari:

## A bag of words you'll come across in the text analytics world

**Algorithm** - The step-by-step process used by a software's engine to determine the meaning/significance of a particular word/phrase.

**Bag-of-Words** - a simplistic way of looking at a text, in which only the words are considered, and their grammar or word order is ignored. In this case, 'man bites dog' = 'dog bites man'.

**Data Mining:** The practice of examining large amounts of data in order to identify themes or new information. Traditionally associated with Structured Data.

**Dictionary Building** - The (often lengthy) human process of defining and categorising words and phrases for your particular area or industry, in order that the software can work with your customers' comments.

**Entity Extraction** - AKA "entity chunking" and "named entity recognition". A process of recognising within a text when specific things are being talked about and allocating them to categories such as place names, personal names, monetary values, products, and expressions of time. It helps the software to 'know' what's being talked about. It requires a dictionary of terms or another source of world-knowledge.

**Machine Learning** - A process by which a software engine "learns" from exposure to pre-labelled sets of training data. In truth, it means you spend time teaching it. Rather than "understanding" language, for example, the software uses a series of statistical equations, counting instances of particular phrases or types.

**Natural Language Processing** - Also called "NLP". A sophisticated model that allows computer systems to accurately assign meaning to different items within a text, for example within a sentence. It is a development from non-NLP sources, which could only count words, without understanding their role in the sentence. It's the "secret sauce" in many of the most advanced text analytics engines, and often relies on machine learning.

**Rule-Based System** - One of two main approaches to text mining, in which some prior knowledge exists within the system (e.g. a knowledge of how language is structured). Compare with statistical or machine-learning systems.

**Sentiment Analysis** - The analysis of a body of text to determine whether it contains positive, negative or mixed opinions about a topic. Because of the difficulty involved in ascribing emotion to any language, sentiment analysis is dogged by problems of inaccuracy. Its utility is marginal at best.

**Structured Data** - Typically 'tick boxes' or where the data is captured in strictly controlled formats, such as date or name, product number or strictly controlled set of possible, discrete responses. Structured data is well-suited to quantitative analysis and can be used for making comparisons, predictions, manipulations etc.

**Text Mining** - Nearly synonymous with "text analytics", though it is sometimes used to refer to the "extraction" phase of a project in which data is categorised and tagged.

**Training** - The process of optimising the software engine for work on a specific text corpus. A major part of the process in any project that doesn't rely on NLP or other machine learning techniques.

**Unstructured Data:** Typically, the freeform text or verbal responses of a customer; but specifically, it's information that is not organized in a pre-defined manner (such as tick boxes). Unstructured information is typically text-heavy, but may contain data such as dates, numbers, and facts as well. Well-suited to being a source of qualitative data to complement structured data.



### Chris West

Chris is Head of Voices at Verbal Identity. He founded the company in 2011.

He is a multiple award-winning copywriter and brand strategy consultant for the world's most interesting brands. He was a post-graduate in Computer Science at Cambridge University.



### Josh Anderson

Josh is an Analyst and Brand Copywriter at Verbal Identity. He joined the company in April.

He is a trained linguist (BA&MPhil, Oxon), with specialisms in linguistic theory and early narratology.



We are committed to expanding everyone's knowledge of what text analytics can do. We are happy to chat about our findings and experiences, or to point you to other sources of information. We undertake text analytics projects with statistical and linguistic interpretation of results for UK and multi-national companies.

To find out more, please call Chris West on +44 (0)20 3053 8329, or email Chris: [YoWesty@verbalidentity.com](mailto:YoWesty@verbalidentity.com)