What is Sentiment Scoring?

Whitepaper

LEXALYTICS
Read ● Between The Lines

Lexalytics on-premise and in-the-cloud sentiment scoring consistently rates the positive or negative assertions that exist within a document or text entity.

Lexalytics on-premise and in-the-cloud sentiment scoring consistently rates the positive or negative assertions that exist within a document or text entity. How to gauge or score the sentiment (sometimes referred to as tone) of a document is a problem originally raised in the context of marketing and business intelligence, where the ability to measure the public's reaction to a new marketing campaign (or a corporate scandal) can have a measurable financial impact on a business.

Historically, measurement of tone was the responsibility of the marketing department and was done by hand. The numerous limitations of this method, such as the manpower required, led to the development of machine scoring.

NEW PRODUCT
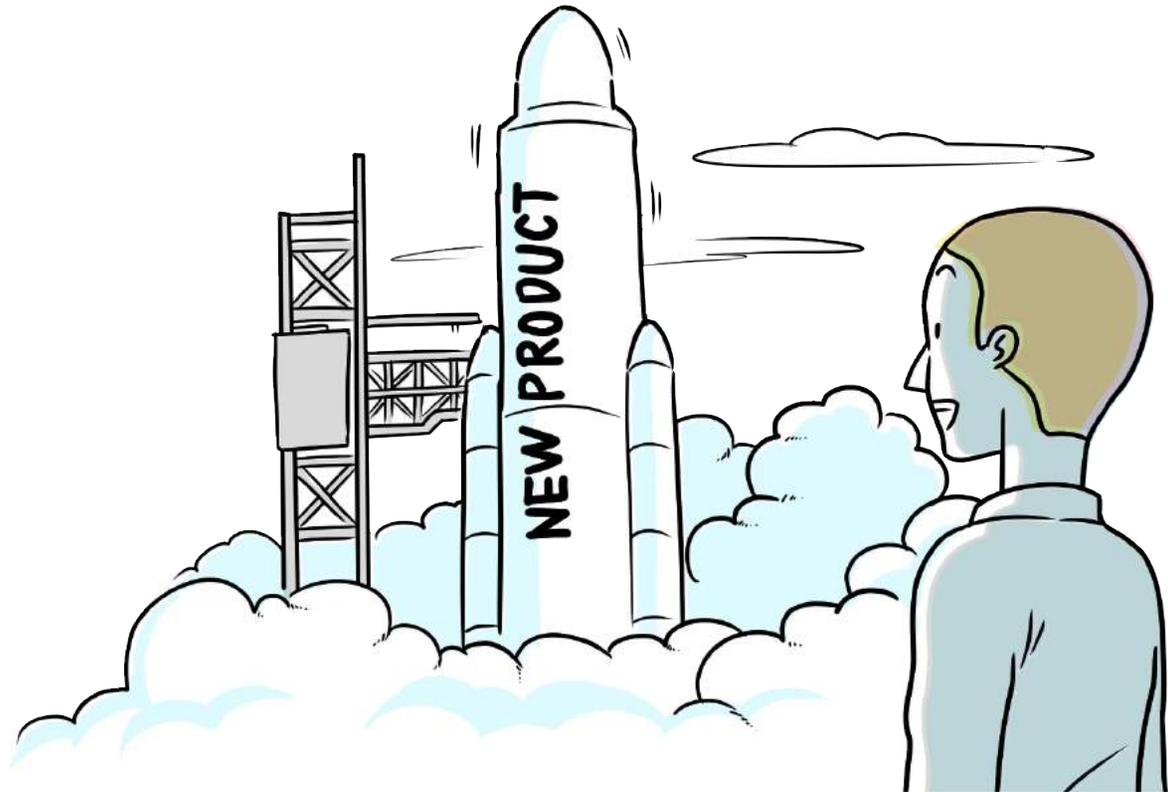
Lexalytics has been perfecting automated analysis of sentiment for over ten years now.

In **financial services**, automated trading algorithms can work in the context of public discussion – similarly, human traders with an ear to a company's public sentiment are better informed in their trading.

Reputation management, the problem every marketing person faces, is much simpler when you have a machine giving you fast, reliable, consistent analyses of the conversations your customers are having about your brand.

Through sentiment analysis, you can discover that you received a wave of negative emails just before a crisis hit — this sort of **eDiscovery** is possible through manual labor, but is far quicker and more reliable with automated analysis.

And as easy as it is to deliver a close-questioned survey to customers, you can't hear the true **voices of your consumers** until you give them the freedom to express their thoughts in their words — on Twitter, Facebook, online review websites, forum comments, and elsewhere.

Reliable sentiment analysis is vital for everything **CEM (customer experience management)** – from the smallest business to the largest political campaign, every organization can benefit from insights into the sentiment of customers, voters, and the public in general — insights that Lexalytics is the best at providing.

Lexalytics has been perfecting automated analysis of sentiment for over ten years now.

As such, we have very real experience in performing and applying sentiment analysis. What's more, Lexalytics supports a multitude of languages:

- English (British and American)
- French (French and Canadian)
- Spanish (Spanish and Mexican)
- Portuguese (Portuguese and Brazilian)
- German
- Mandarin (Traditional and Simplified)
- Korean
- Italian
- Japanese
- Malay and Singlish
- Dutch
- Arabic*
- Russian*

*Does not support phrase-based sentiment, only entity sentiment
Few vendors have been in the business as long as we have, and no other sentiment solution offers the same level of functionality and sophistication.

The following whitepaper will demystify sentiment scoring and explain how the Lexalytics sentiment system works. This includes a discussion of how and why we have extended the basic concept of document sentiment to the theme, topic, and entity level, and how this technology is being further extended to measure other indicators within content, such as the assessment of threat, customer satisfaction and many other contextual indicators.

Humans are very good at reading a sentence and mentally scoring the sentiment (we average around 80% accuracy), but historically computers struggled to match a human's accuracy — until Lexalytics came along.

Consider these sentences:

- A horrible pitching performance resulted in another devastating loss.
- Sub-par pitching and superb hitting combined to cost us another close game.

They both have the same basic topic, the loss of a baseball game, but obviously (to you!) the first sentence is contextually much more negative. To discern this, humans focus on the emotive phrases: "horrible pitching" and "devastating loss". **The sentiment system developed by Lexalytics does exactly the same thing.**

Our software identifies the emotive phrases within a document, scores these phrases (roughly - 1 to +1), and then combines these scores to discern the overall sentiment of the sentence.

Importantly, our sentiment scoring will score sentences the same way every time the engine is exposed to them – our software is not affected by whether or not it's had its morning coffee, or whether it's a fan of the team that lost (or that won!). That's consistency, and that's important.

The first step in determining the tone of a document is to break it into its basic parts of speech (POS tagging).

POS tagging is a mature technology that identifies all the structural elements, including verbs, nouns, adjectives, and adverbs.

In most cases, sentiment is expressed in adjective-noun combinations like "horrible pitching" and "devastating loss". **That's how our software works, too.** We call these combinations **"sentiment-bearing phrases"**

Our software assigns a number to the sentiment: what we've done at Lexalytics is create a very, very large dictionary of sentiment-bearing phrases and their relative scores. These scores are pre-determined by how frequently a given phrase occurs near a set of known good words (e.g. good, wonderful, spectacular) and a set of bad words (e.g. bad, horrible, awful). In addition, over the years we've extensively tuned this dictionary for better accuracy.

We used an extremely large corpus of text (the web, via an internet search engine) to evaluate the nearness of known good and bad words to the phrase being considered. Consider the phrase "devastating loss". It means something to you because you've come to associate those words with something bad. That's exactly the process that we go through – we check to see if we should associate a phrase with positive or negative sentiment, and just how closely we should associate it.

Here's how this would work in a document:

Dan Wells joins Nautilus Footwear, America's fastest growing safety shoe company, as Vice President Nursing Division. "We have ended 2000 with record sales and earnings. Phase 3 of our strategic plan is

to clearly separate our three business units of industrial, nursing and public safety. Dan will lead our nursing strategy and drive the efforts of our nursing sales team. The clear separation of our nursing business is a natural for us as we have been a leader in the nursing arena since beginning the company in 1996," said Wayne Easley, President / CEO. "Dan comes to us with an extensive background in the footwear industry most recently with Berkshire Hathaway's, Lowell Shoe Company. He has solid account and product knowledge that can clearly continue to deliver our 'ergonomic message' to the Nursing community," added Elsey.

The green phrases below are the sentiment bearing phrases in the document (these happen to all be positive sentiment bearing phrases):

[Figure 1 Whole Document Sentiment Scoring]

We have an algorithm that we use to combine phrase scores in the document based on an operation called lexical chaining. The overall document sentiment for the above text comes out to 0.365 (solidly positive)."

Unfortunately, except for press releases, it's rare to find a document that is homogeneously positive or negative.

Sentiment is typically a localized phenomenon that is more appropriately computed at the paragraph, sentence or entity level. Consider the following example:

"Julie Jones superb performance in the gubernatorial debate has all but assured her of victory in the upcoming elections. Unfortunately, the evening did not go as well for her opponent John Adams. Mr. Adams nervous and uncertain performance has put his entire political future into question."

The sentiment of this document is completely different for the two individuals described within, while the overall sentiment for the document averages out to roughly neutral.

Let's look at the results of this snippet at the overall level and at the entity level (red are negative sentiment bearing phrases, green are positive):
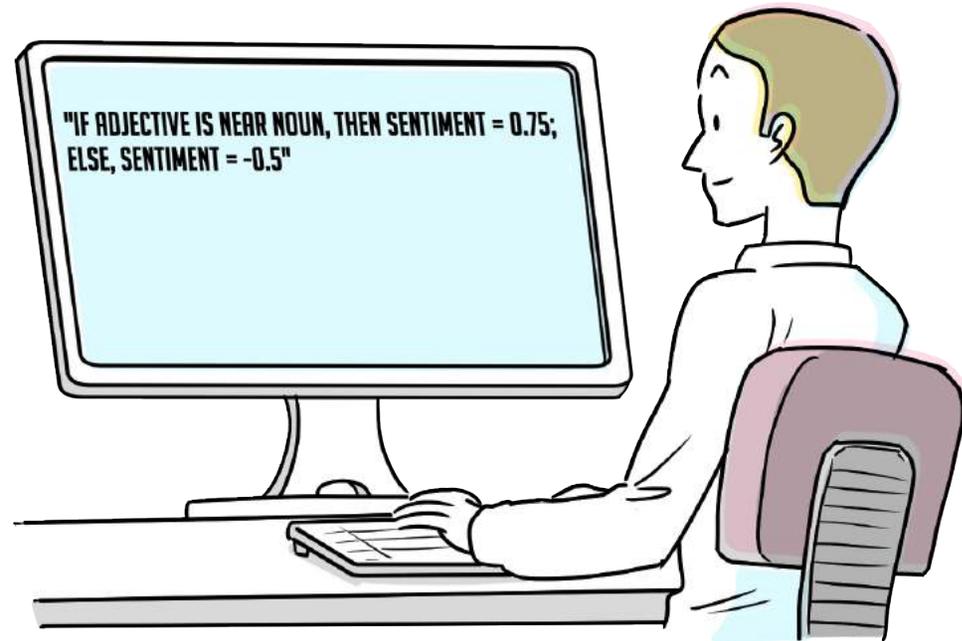
"IF ADJECTIVE IS NEAR NOUN, THEN SENTIMENT = 0.75; ELSE, SENTIMENT = -0.5"

SENTIMENT TAGGED TEXT

Julie Jones superb performance in the gubernatorial debate has all but assured her of a major victory in the upcoming elections. Unfortunately, the evening did not go as well for her opponent John Adams. Mr. Adams nervous and uncertain performance has all but guaranteed a loss and put his entire political future into question.

Now let's examine the entity level sentiments for each person, and start by identifying every instance where the person is mentioned. How we determine what constitutes an entity is beyond the scope of this whitepaper, but isn't important for understanding how we assign sentiment:

ENTITY TAGGED TEXT

Julie Jones superb performance in the gubernatorial debate has all but assured her of a major victory in the upcoming elections. Unfortunately, the evening did not go as well for her opponent John Adams. Mr. Adams nervous and uncertain performance has all but guaranteed a loss and put his entire political future into question.

In this tagged text, notice that the system identifies not only the people by name, but also identifies the pronouns "her" and "his" (and will correctly associate the "her" with Julie Jones – an operation called pronominal co-referencing). Correctly including the pronouns significantly improves our software's ability to measure the tone for each individual. Computing sentiment for each entity in this block yields:

- Julie Jones: Positive (+) 0.22
- John Adams: Negative (-) 0.11

Our ability to accurately and reliably associate sentiment back to the relevant entity sets us apart from other providers that measure sentiment. Refined processes like the Lexalytics **Concept Matrix™**, **lexical chaining**, and **pronominal coreferencing** (identifying when two or more pronouns refer to the same thing) enable our software to focus on the sentiment or tone of specific people, companies or products. Consider the case of a product review article, where one product gets panned and the other doesn't – if you can't discern which was which, the document doesn't do you any good. The true value of measuring sentiment is in applying the measure to the people or products you're concerned about. Lexalytics sentiment analytics give you results that you can trust and take action on.

Another important piece of sentiment analysis is the idea of negation: when one part of a sentence cancels out or flips the sentiment of the rest. Consider the following sentences:

- "It's been a good day."
- "It's not been a good day."
- "It was not only a good day but a productive day as well."

The first sentence is strongly positive and unambiguous. But one added word in the second sentence, and the meaning – and sentiment – are flipped. That extra word, "not", negates the positive sentiment of "good".
The third sentence is tricky. There are elements of negation, but they aren't so cut-and-dry as the second sentence. Lexalytics' revolutionary **Syntax Matrix™**, an innovative new way to understand sentence structure, aids in the analysis of this sentence's sentiment: the Syntax Matrix™ identifies "but" as a modifier, in this case adding more positive sentiment. Given a sentence like,

"The food was good but the service was awful."

the Matrix identifies "but" as dividing the sentence into two halves with very different sentiment, and evaluates each piece separately. Additionally, the Matrix knows to weigh the more strongly sentimental piece, the second chunk, more heavily. Sentiment analysis software as refined as ours will pick up on these subtleties and adjusts its evaluation accordingly – just another of the myriad pieces that make Lexalytics' sentiment analysis the best in the world.
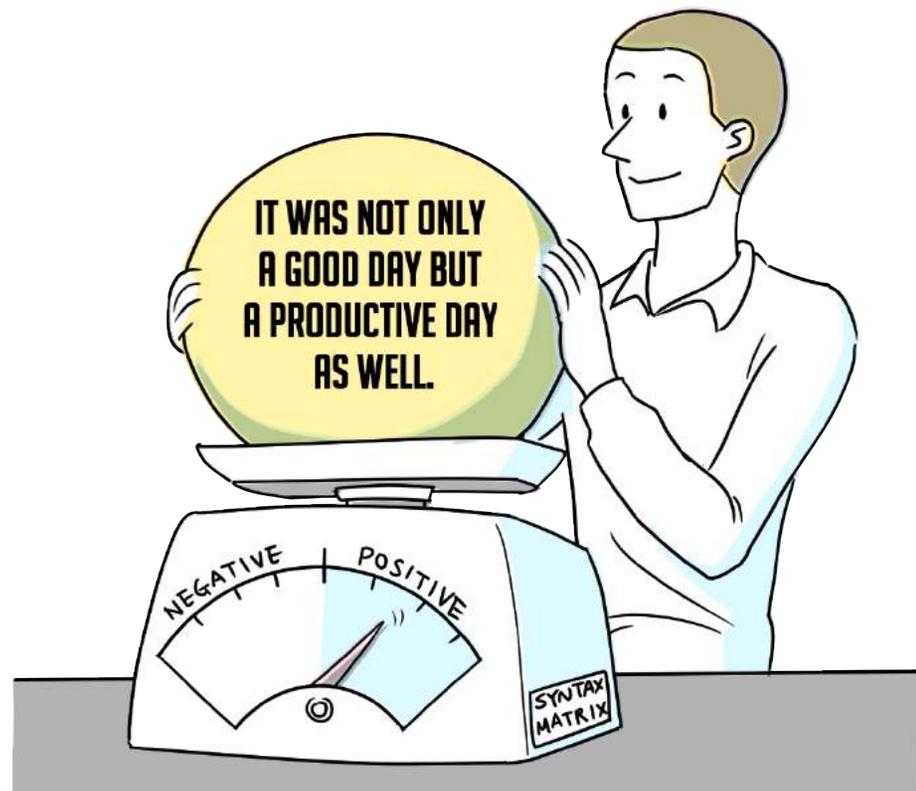
But wait, we're not done yet.

With the rise of short-form social platforms like Twitter, social media conversations are awash with hashtags, acronyms, and composite words – to properly monitor conversations about your brand, you must be able to interpret this "Internet speak". Lucky for you, Lexalytics supports all of the above.

We were the first to offer support for **emoticons**, and maintain a substantial library of emoticons and their associated sentiment: we can tell the difference between :-P and :-).

WTF? FTW. LOL! No matter the **acronym**, we know and analyze it.

Hashtags were introduced with Twitter and quickly spread to the rest of social media. Hashtag phrases can be critical for sentiment evaluation, and we offer the only sentiment analysis solutions that can expand hashtags into their component words (#ilovelexalytics becomes "I love Lexalytics") for use by our sentiment and entity extraction algorithms.

And while just about everything gets made into an emoji these days, ours is the only system that can analyze the sentiment of emoji. We even categorize emoji: a picture of a piece of pizza gets tagged as food, for example. Emoji can carry a huge sentiment weight compressed into a single character, and lesser products skip over that sentiment.
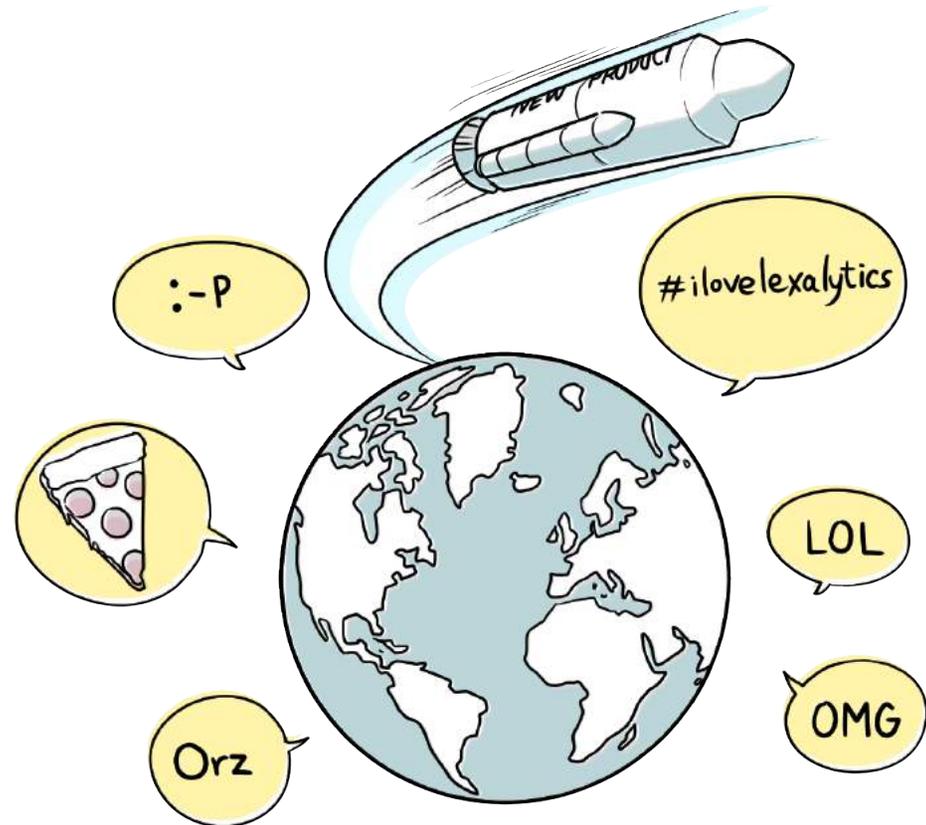
The determination of sentiment is another step in the process of turning unstructured content into structured data, so that the humans who interact with this ever-expanding sea of information can spot trends and patterns within. These trends and patterns inform people in decision-making for a host of industries, including financial services, reputation management, eDiscovery, voice of customer, customer experience management, and more.

Lexalytics® is the industry leader in translating text into profitable decisions. Lexalytics deploys state-of-the-art on-premise and in-the-cloud text and sentiment analysis technologies that process billions of unstructured documents every day globally, transforming customers' thoughts and conversations into actionable insights. The on-premise Salience® and SaaS Semantria® platforms are implemented in a variety of industries for social media monitoring, reputation management and voice of the customer programs.

Lexalytics is based in Boston, MA, and has offices in the U.S. and Canada. For more information, please visit www.lexalytics.com, email sales@lexalytics.com or call 1-617-249-1049. Follow Lexalytics on Twitter, Facebook, and LinkedIn for updates and insights into the world of text mining.

L E X A L Y T I C S

Read ● Between The Lines
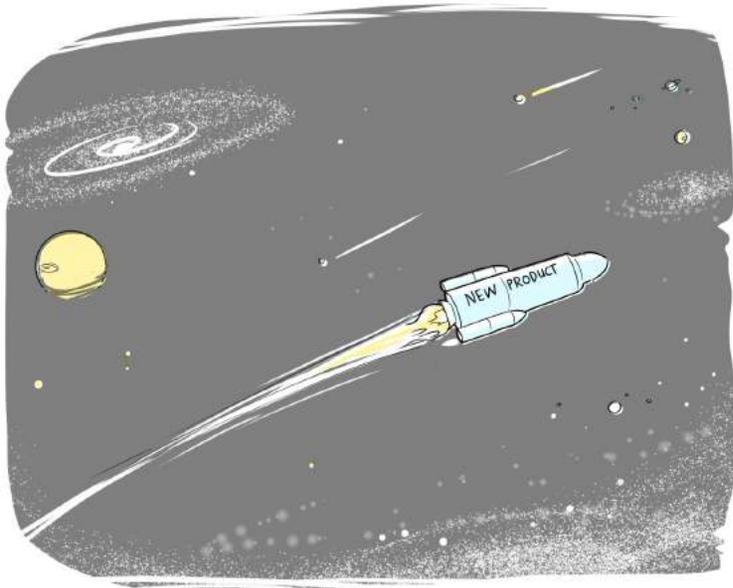
320 Congress St
Boston, MA 02210

General Inquiries
1-800-377-8036

Sales
sales @lexalytics.com
1-800-377-8036 x1

International
1-617-249-1049