

What is Sentiment?

Sentiment is a Luminoso Daylight[®] feature that examines a set of documents and reports the presence of sentiment – positive, negative, or neutral.

Luminoso analyzes text data at the document level by examining the sentiment of words and phrases. When first entering the feature view, a summary is provided to help users get started. Sentiment currently supports 13 languages, including Arabic, Chinese, Dutch, English, French, German, Italian, Japanese, Korean, Portuguese, Russian, Spanish, and Swedish, including emoji and emoticons. Development of sentiment lists for Polish and Bahasa Indonesia is underway and expected to be released in early Q3 2019.

Using Sentiment in Luminoso Daylight

Sentiment words

To assemble the various views of sentiment in a project, Luminoso searches for and analyzes sentiment words, or words that convey feeling and indicate sentiment around topics. These words are often adjectives. Consider the following document:

"The food was superb, but my waiter was slow"

In this example, "superb" and "slow" are considered sentiment words, as they indicate feeling around the food and the waitstaff, respectively.

Sentiment score: Document-level analysis

When a project is created, Daylight calculates a sentiment score for every document by assigning any sentiment-laden word a sentiment integer rating between -5 and 5. The ultimate distribution of this integer is reflected in the app as a number from -1.0 to 1.0 and scored as "negative" for anything below 0, and "positive" for anything above 0. If a document's sentiment score is closer to -1, the overall document is likely negative. If the sentiment score is closer to 1, the document is likely positive. When a project is created, Luminoso calculates a sentiment score for each document.

To calculate each document's sentiment score, Luminoso sums individual scores for each sentiment word at the concept level, then converts that sum into a percentage.

4 -1 "The food was <mark>superb</mark>, but my waiter was <mark>slow</mark>"

In the above example, "superb" received an individual sentiment score of 4, and "slow" a score of -1. As the positive score is much stronger than the negative, the entire document will have a positive sentiment of 3, expressed as the percentage 30%. The system interprets this overall document as having a 30% chance of being positive.

Sentiment mix: Concept-level analysis

Sentiment mix is a measure of all positive, negative, and neutral sentiment about topics in a dataset, described using three representative percentage values. Expressed as a combination, this provides a unique view into the mix of feelings around a particular word or phrase. Understanding a concept's sentiment mix is extremely valuable when analyzing datasets that contain no ratings or have a statistically insignificant number of rating responses.

Sentiment mix is calculated by determining each document's sentiment score, sorting by overall positive, negative, and neutral documents, summing the number of each positive, negative, and neutral document, and then calculating the percentage of each sentiment type over the total number of documents in the dataset.

Consider a group of 20 beer review documents, a subset of which are represented below. The phrase "dark chocolate" appears in these following four:

"The aroma is massively roasty with lots of black malts, cocoa powder, **dark chocolate** and espresso."

"Big **dark chocolate** flavor, roasted malt, freshly brewed coffee, nice hint of bourbon, and an excellent vanilla extract taste."

"Pours pitch black with a two-finger **dark chocolate**/coffee-colored head with **excellent** retention, only **slowly** fading into a lasting cap that coats the glass with chunky rings of soapy lacing."

"Starts off with a hoppy and roasted coffee bean kick that melts away into a sweet dark chocolate and malty finish that hangs on the back of the palate, which is wonderful."

When uploaded, the application searches for sentiment words, analyzes their usage, and determines a sentiment score for each individual document. Based on the subset of documents in which "dark chocolate" appears, there are 0 negative, 1 neutral, and 3 positive associated documents. Translated into a percentage, calculated over the total set of 20 documents, the resultant sentiment mix for "dark chocolate" would be assigned a score of 0% / 5% / 15%, or 0% negative, 5% neutral, and 15% positive.

Sentiment suggestions

In the Sentiment feature pane, Luminoso displays a list of the 50 concepts most significantly correlated to sentiment in each project. This list of concepts contains topics, phrases, and nouns that are associated with strong sentiment, and not actual sentiment words, such as "great", "amazing", or "awful", which are inherently descriptors and used to calculate scores in the feature.

Frequently asked questions

Why is neutral sentiment for concepts not displayed by default?

As Sentiment is designed to show concepts with the most positive and negative sentiment, neutral results were found to be uninteresting in comparison. Neutral sentiment can be viewed in the application by either hovering over a concept result or inclusion in the results export.

What are the current limitations of sentiment?

Very large documents. Luminoso analyzes projects at the document level, meaning if documents contain multiple sentiment words, some sentiment terms may get buried under others. For example, consider the previous document:

4 -1 "The food was <mark>superb</mark>, but my waiter was <mark>slow</mark>"

This document has an overall positive sentiment. If multiple documents reiterate both a highly positive word such as "superb", in conjunction with negative feedback about the waitstaff, it is possible for the sentiment surrounding the waitstaff to get buried. This type of issue manifests in large datasets such as those examining Voice of the Employee, where respondents wish to convey a general feeling of positivity around their work environments, and only a bit of criticism. The result? Positive terms mask much less frequent negative terms. This problem is usually mitigated by feature- or aspect-based sentiment, which works by assigning a sentiment score to each individual feature/aspect, not document. Feature-based sentiment is currently available in Luminoso as a solution engagement.

Wishful thinking and sarcasm. Luminoso cannot detect nuances such as wishful thinking or sarcasm, where positive or negative sentiment is used to convey its opposite. For example, consider the following document:

"If only the company offered a generous tuition reimbursement or a student loan assistance benefit. It would be sooo great to have assistance on student loan repayment. This would increase the level of talent and attract amazing candidates."

This document would be scored highly positive, even though it's clear to a human reader that the company lacks this benefit. Sentiment classification is currently unable to differentiate this tone from a direct answer.

Social nuances. Sentiment has no awareness of social nuances or norms. Consider this mobile gaming review document:

"I absolutely love this game. Only one problem. Men seem to think it is a dating site. Have to be careful about that."

This document would be assigned a positive sentiment score, even though it expresses negative criticism that using the game's chat functionality as a dating service is discouraging.