

Amazon Comprehend

Amazon Comprehend is an Amazon service for Natural Language Processing which helps to find relationships and insights in text. It lets you integrate powerful natural language processing into your apps with a simple API call. The service identifies the language of the text; extracts key phrases, places, people, brands, or events; understands how positive or negative the text is; analyzes text using tokenization and parts of speech; and automatically organizes a collection of text files by topic. You can also use AutoML capabilities in Amazon Comprehend to build a custom set of entities or text classification models that are tailored uniquely to your organization's needs.

Main capabilities

Sentiment

Allows you to understand what the user is saying if it is positive or negative.

Entities

These features extract the entities and categorizes them for you. Entities can be anything like objects, people, organizations.

Languages

This feature detects the language in which the text is written or presented.

Key phrases

It is like parts of speech tagging finding what is the word if it is a noun, verb, adjective.

Topic Modeling

Extracts up to 100 topics from a corpus of documents and helps you to organize the documents into the data.

Benefits

Get better answers from your text

Amazon Comprehend can discover the meaning and relationships in text from customer support incidents, product reviews, social media feeds, news articles, documents, and other sources. For example, you can identify the feature that is most often mentioned when customers are happy or unhappy about your product.

Organize documents by topics

Amazon Comprehend can analyze a collection of documents and other text files (such as social media posts) and automatically organize them by relevant terms or topics. You can then use the topics to deliver personalized content to your customers or to provide richer search and navigation. For example, if you have an extensive collection of news articles, you can automatically group them by subject matter to enable your site to suggest new articles to visitors based on what they have read previously.

Train models on your own data

You can easily extend Amazon Comprehend to identify specific terms, such as policy numbers or part codes. You can also extend Comprehend to classify documents and messages in a way that makes sense for your business, like customer support inquiries by request or social media posts by product. Adding this customization requires no machine learning expertise. You simply provide your labels and a small set of examples for each and Comprehend takes care of the rest.

Support for general and industry specific text

Powered by state-of-the-art machine learning models, Amazon Comprehend can discover insights from unstructured text like social media posts, emails, and web pages. Amazon Comprehend Medical also identifies medical information, such as medication and medical conditions, and determines their relationship to each other (e.g., medicine dosage and strength). For example, Amazon Comprehend Medical extracts “methicillin-resistant Staphylococcus aureus,” often inputted as “MRSA,” and provides context, such as whether a patient has tested positive or negative, to make the extracted term meaningful.

How it works

Amazon Comprehend works in 3 steps

It takes any unstructured text like Social media posts, documents, emails, web pages, phone transcripts, medical records as input

Amazon Comprehend under the hood use the text classification models and extracts the key phrases, entities, sentiment language, syntax, topics, and document Classifications

The final output shows the entities, topics, and document classifications with the confidence scores.

References

<https://aws.amazon.com/comprehend/>

<https://blog.usejournal.com/sentiment-analysis-using-amazon-comprehend-one-of-the-tools-for-natural-language-processing-7d16c41d51ca>