What is a Text Annotation Tool?

Text annotation is the machine learning process of assigning labels to a text document or different elements of its content to identify the characteristics of sentences. As intelligent as machines can get, human language is sometimes hard to decode, even for humans. In text annotation, sentence components, or structures are highlighted by certain criteria to prepare datasets to train a model that can effectively recognize the human language, intent, or emotion behind the words. The training data is given to machine learning so they can comprehend various aspects of sentence formation and conversations between humans.

Why is it important?

You might still wonder; why do we need to annotate text at all? Recent breakthroughs in NLP highlighted the escalating need for textual data for applications as diverse as insurance, healthcare, banking, telecom, and so on. Text annotation is crucial as it makes sure that the target reader, in this case, the machine learning (ML) model, can perceive and draw insights based on the information provided. As the world becomes more digitized, data quality needs also increase rapidly. Businesses must learn how to get the best use of the large amounts of data that are provided to their platforms to stand out in the market. Not to mention the increasing demand of customers for digitized and timely support services. We'll take a deeper dive into particular use cases later in this post, but for now, keep the following in mind: textual data is still data—much like images or videos—and is similarly used for training and testing purposes.

How is text annotated: NLP text annotation

The list of tasks computers are taught to perform increases steadily, yet some activities still remain untackled: natural language processing (NLP) is no exception to that. Without human annotators, models won't acquire the depth, nativity, and even slang in which humans craft, control, and manipulate language. That's why companies continuously turn to human annotators to ensure sufficient amounts of quality training data. Current NLP-based artificial intelligence (AI) solutions cover voice assistants, machine translators, smart chatbots, and alternative search engines, yet the list keeps expanding in parallel with the flexibility text annotation types propose.

Example
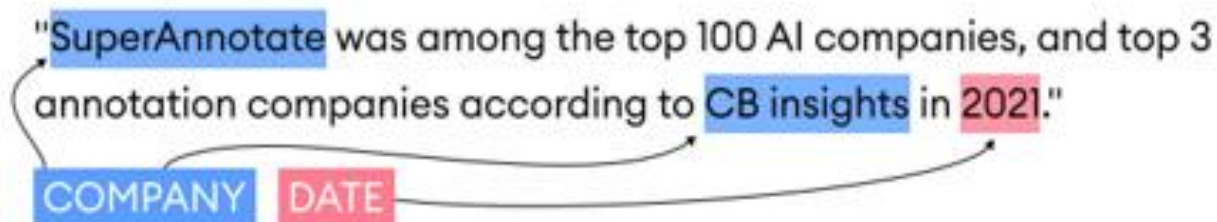
Text Annotation Tools

The following table lists various text annotation tools, their collaborative capabilities, pricing, and key features:

Tool Name	Collaborative	Pricing	Features
Label Studio	Yes	Free & Paid	Supports entity recognition, sentiment analysis, text classification, and more.
Prodigy	Limited	Paid	Optimized for NLP tasks with pre-trained models; customizable.
Doccano	Yes	Free	Supports text classification, sequence labeling, and sequence-to-sequence tasks.
Amazon SageMaker Ground Truth	Yes	Paid	AI-assisted text labeling; integrates with AWS services.
LightTag	Yes	Paid	Team collaboration, auto-suggestions, workflow management.
SuperAnnotate	Yes	Paid (Trial)	Annotation for text, images, and video; integrated QA features.
Tagtog	Yes	Free & Paid	Custom annotation schemas; suitable for NLP and medical text.
INCEpTION	Yes	Free	Team annotation, versioning, and evaluation.
Dataturks	Yes	Free & Paid	User-friendly interface for NLP tasks.
brat	Limited	Free	Browser-based, simple for text annotation tasks.
UBIAI	Yes	Paid (Trial)	User-friendly with pre-annotations and team collaboration.
MonkeyLearn	Yes	Free & Paid	Designed for text classification and sentiment analysis.
Snorkel	Yes	Paid	Programmatic labeling with weak supervision.