
Machine Learning / AI and NLP with Hubly

Overview

Obtain and comprehend insightful information from text stored in digital formats and documents

- Discover useful insights from text in documents, emails, social media feeds, support queries, product reviews, and more.
- Extracting text, important phrases, subjects, sentiment, and more from documents like insurance claims can simplify document processing operations.
- Create a differentiator for your company by teaching a model to categorise documents and recognise terms without any prior knowledge of machine learning.
- By locating and removing Personally Identifiable Information (PII) from documents, you may safeguard and regulate who has access to your sensitive data.

What is Natural Language Processing?

Natural language processing (NLP) is the ability of a computer program to understand human language as it is spoken and written and is known as natural language processing (NLP). It is a component of artificial intelligence (AI)

How does natural language processing work?

NLP enables computers to understand natural language as humans do. Whether the language is spoken or written, natural language processing uses artificial intelligence to take real-world input, process it, and make sense of it in a way a computer can understand. Just as humans have different sensors -- such as ears to hear and eyes to see - computers have programs to read and microphones to collect audio. And just as humans have a brain to process that input, computers have a program to process their respective inputs. At some point in processing, the input is converted to code that the computer can understand.

How does natural language processing work?

NLP makes it possible for computers to comprehend natural language just like people do. Natural language processing employs artificial intelligence to translate real-world information into a form that a computer can comprehend, whether the language is spoken or written. Computers have programmes for reading and microphones for audio collection, much as humans have various sensors, such as ears for hearing and eyes for seeing. Computers have programmes to process their different inputs, just as people have brains to process that input. The input is eventually translated into code that the computer can comprehend during processing.

Hubshield AI - Natural Language Processing

Hubshield Lucid is a natural language processing (NLP) service via the Hubshield Content Moderation AI that extracts information from text using machine learning.

You can quickly incorporate natural language processing into your apps with the help of Hubshield Lucid's Custom Entity Recognition, Custom Classification, Key phrase Extraction, Sentiment Analysis, Entity Recognition, and other APIs.

Simply give the location of the source text or document and contact the Hubshield Lucid APIs from within your application. You can use the entities, key phrases, emotion, and language that the APIs output in your application in JSON format.

Personalised Entity Recognition

You can modify Hubshield Lucid using Custom Entity Recognition to find words that are particular to your organisation. Without the need for machine learning, Hubshield Lucid will use AutoML to learn from a small set of examples (such as a list of policy numbers, claim numbers, or SSNs) and then train a private, custom model to recognise terms like claim numbers in any other block of text within PDFs, plain text, or Microsoft Word documents.

Text Example:

Hello, my name is Jack Jones, and I'm here to make a vehicle accident claim. My policy number is 456-YQT

Entity	Category	Count	Confidence
--------	----------	-------	------------

“If you could help prevent online abuse. Would you do it?”



456-YQT	Policy ID	1	0.95
---------	-----------	---	------

Individual Classification

Using your company-specific labels and the Unique Classification API, you can quickly create custom text classification models without having to master machine learning. For instance, your customer care organisation can utilise Custom Classification to automatically categorise incoming requests by problem type based on how the customer has characterised the issue. Moderating website comments, prioritising client input, and organising workgroup documents are all made simple with your bespoke model.

Text	Label	Confidence Score
Line 0	Account Question	0.95
Line 1	Ticket Refund	1
Line 2	Flight Complaint	1
Line 3	Flight Complaint	0.93
Line 5	Ticket Refund	1

Recognition of Entity

The named entities ("People," "Places," "Locations," etc.) that are automatically categorised based on the input text are returned by the Entity Recognition API.

Sample Text: Hubly is located in London, UK and was founded by Jack Jones in 2019 enabling customers to quickly develop engaging online communities and protect their members from Toxic and Abusive online content. London is the capital of England. Other companies based in London are Lloyds Insurance and Hubshield

Entity	Category	Confidence
Hubly	Organisation	0.96

London, UK	Location	0.95
2019	Date	0.99
Jack Jones	Person	0.98
England	Location	0.98
Capital of England	Location	0.97
Lloyds Insurance	Organisation	0.99
Hubshield	Organisation	0.99

Sentiment Analysis

The overall sentiment of a text is returned by the Sentiment Analysis API (Positive, Negative, Neutral, or Mixed).

Example: I ordered a small and thought it would fit well, but it was more like a medium-large. It had excellent quality. It's a little lighter than in the picture, but it's close. It would be ten times better if the interior was lined with cotton or wool.

Sentiment	Score
Mixed	0.87
Positive	0.7
Negative	0.2
Neutral	0.00

Targeted Sentiment

By determining the sentiment (positive, negative, neutral, or mixed) towards things within the text, Targeted Sentiment offers more detailed sentiment information.

Sample Text: The service was slow, but I loved the burger.

Text	Entity Type	Confidence score	Sentiment	Score
I	Person	0.99	Neutral	0.99
Burger	Other	0.99	Positive	0.99
Service	Attribute	0.99	Negative	0.99

Redaction and Personal identifiable information

To find and remove personally identifiable information (PII) from customer emails, support tickets, product reviews, social media posts, and other sources, use Hubshield Lucid ML capabilities. ML knowledge is not necessary. For instance, before indexing the documents in the search solution, you can scan support requests and knowledge articles to find PII entities and redact the text. After then, PII entities in documents are not present in search solutions. You can protect your privacy and follow local rules and regulations by redacting PII entities.

Example Text: Hello, this is Jack Jones. The minimum payment due on your ABCCompany Financial Services Ltd credit card account 0000-0000-0000-1234 is £30.00 and is due on June 30th 2023. Your payment will be taken on the due date from your bank account number XXXXXX1234 with the routing number XXXXX0000 in accordance with your autopay preferences.

Entity	Type	Score
Jack Jones	Name	0.99+
0000-0000-0000-1234	Debit Card Nubmer	0.99+
June 30th	Date / Time	0.99+
XXXXXX1234	Bank Account Number	0.99+
XXXXXX0000	Bank Routing Number	0.99+

Keyphrase Extraction

The talking points or key phrases are returned by the Keyphrase Extraction API along with a confidence score indicating whether or not this is an important phrase.

Example: A customer compares an Instant Camera to a DSLR. The API extracts key phrases and returns a confidence score about the results

Text example: I love taking pictures, and I usually use my DSLR or a disposable instant camera that I keep with me for casual photography. Although my DSLR is superior in terms of power and practicality, my instant film camera has a certain quality. Maybe it's because you're using real film, or maybe it's because each photograph you take is a special physical artefact (which is noteworthy in the age of Instagram and Facebook, where pictures are many). All I can say for sure is that they are a tonne of fun to use and that when you bring one of these out at a party, everyone's eyes light up.

Keyphrase	Confidence
i love taking pictures	0.99
my DSLR	0.97
disposable instant camera	0.99
casual photography	0.99
power and practicality	0.94
real film	0.99
each photograph	0.92
special physical artefact	0.99
Age	0.91
of	0.99
Instagram and Facebook	0.99

Events Detection

With Hubshield Lucid Events, you can extract the event structure from a document, turning pages of text into quickly processed data that can be used by graph visualisation tools or AI applications. With this API, you can quickly and efficiently respond to who-what-when-where queries across huge document sets without any prior NLP expertise. To extract specific information from unstructured text about actual events and related entities, use Hubshield Lucid Events.

Detection of languages

The Language Detection API recognises text that has been written in more than 100 different languages automatically and delivers the language that is most prevalent, along with a confidence score to substantiate that claim. For further information, see this [documentation page](#).

Below we search for the

Example of Text: Hubly Elastic Compute Cloud (Amazon EC2) è un servizio Web che fornisce capacità di elaborazione sicura e scalabile nel cloud. È concepito per rendere più semplice il cloud computing su scala Web per gli sviluppatori.

ISO-639-1 Language Code	Language	Confidence
it	italian	1.0

Syntax Analysis

Customers can analyse text using tokenization and Parts of Speech (PoS) and recognise word boundaries and labels like nouns and adjectives using the Hubshield Lucid Syntax API.

Text example: I adore my quick, new Ipad Pro!

Text	Tag
I	Pronoun
adore	Verb
my	Pronoun

quick	Adjective
,	Punctuation
new	Adjective
Ipad	Proper noun
Pro	Proper noun
!	Punctuation

Topic Modeling

From a group of documents kept on the database, Topic Modeling extracts pertinent terms or subjects. It will determine which themes are covered the most frequently in the collection, group them, and then map which documents fall under each topic.

Hubshield AI can analyse the docs and return two views

1. Combining related keywords into groups.

Each keyword cluster belongs to a topic group. Weight reflects how often a keyword is within the group. The context of the topic group is best indicated by keywords with weights that are near to 1.

Topic Group	Keywords	Weight
1	Hubly	0.83
1	London	0.71
2	Flights	0.74
2	Offices	0.64

Each keyword cluster belongs to a topic group. Weight reflects how often a keyword is within the group. The context of the topic group is best indicated by keywords with weights that are near to 1.

2. Topic-based grouping of documents

Document Name	Topic Group	Proportion
Doc1.txt	1	0.83
Doc2.txt	1	0.71
Doc3.txt	2	0.74
Doc3.txt	2	0.64

Based on the percentage of the topic group's weighted keywords that are present in each document, each one is assigned to a subject group.

Support for multiple languages

Text in German, English, Spanish, Italian, Portuguese, French, Japanese, Korean, Hindi, Arabic, Chinese (simplified), Chinese (traditional), and more may all be analysed using Hubshield Lucid .

Customers can utilise Hubly Translate to translate text into a language that Hubshield Lucid supports before using Hubshield Lucid to carry out text analysis in order to build apps in additional languages. Refer to the documentation page for further information on language support.