Google Cloud Platform Natural Language Processing API

Google Machine learning APIs:

* Cloud Machine Learning Engine
* Vision API
* Speech API
* Natural Language API
* Prediction API (also provides sentiment analysis feature, soon to be deprecated in april 2018)
* Translation API

Exploring more into Natural Language API 🡪

Programming language supported:

|  |  |  |
| --- | --- | --- |
| Go | Java | .NET |
| Node.js | PHP | Python |
| Ruby |  |  |

Natural Language API supports following features [output – JSON]:

* **Sentiment analysis** - Extract tokens and sentences, identify parts of speech (PoS) and create dependency parse trees for each sentence.
* **Entity analysis** - Identify entities and label by types such as person, organization, location, events, products and media.
* **Entity sentiment analysis** - Understand the sentiment for each mention of an entity within a block of text.
* **Syntax analysis** - Understand the overall sentiment expressed in a block of text.
* **Multi\_language** - Enables you to easily analyze text in multiple languages including English, Spanish, Japanese, Chinese (Simplified and Traditional), French, German, Italian, Korean and Portuguese.
* **Integrated REST API** - Access via REST API. Text can be uploaded in the request or integrated with Google cloud storage

Language Support

|  |  |  |
| --- | --- | --- |
| English | Chinese | French |
| German | Italian | Japanese |
| Korean | Portuguese | Spanish |

The API supports above languages which can be specified using ‘Language’ parameter that is optional. If this parameter is not specified, then the language is auto detected by NL API.

Can be deployed using

* Client libraries
* REST APIs

Pricing:

Cost of API-

The Cloud Natural Language API is priced using units of measurement known as text records. A text record may contain up to 1,000 Unicode characters within the text content sent to the API for evaluation. Text in excess of these 1,000 characters counts as additional record(s). Prices are expressed in dollars per 1,000 text records.

Prices for usage of the Cloud Natural Language API are computed monthly based on which feature of the API is used, and how many text records are evaluated using those features. These prices are noted in the table below.

Monthly prices (per 1,000 text records)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Feature | 0 - 5K | 5K+ - 1M | 1M+ - 5M | 5M+ - 20M |
| **Entity Analysis** | Free | $1.00 | $0.50 | $0.25 |
| **Sentiment Analysis** | Free | $1.00 | $0.50 | $0.25 |
| **Syntax Analysis** | Free | $0.50 | $0.25 | $0.125 |
| **Entity Sentiment Analysis (beta)** | Free | $2.00 | $1.00 | $0.50 |

Cost of platform-

We may also be charged for Google cloud platform resources used in our project such as Google computer engine instances, google cloud storage etc.

Google has provided a [Cost calculator](https://cloud.google.com/products/calculator/)

Model Deployment-

We can host our trained models in the cloud with Cloud ML Engine and use them to get predictions. Need to store model in google cloud storage

You can get a model by:

* Following the Cloud ML Engine training steps to train in the cloud.
* Training elsewhere and exporting to a saved model.

Then we need to create a model version

We can then use our deployed model version to get Online or batch predictions.

Online-

Request online predictions by sending your input data instances, in the required format in a JSON string, to projects.predict.

Batch-

When we don't need your predictions right away, or when we have a large number of instances to get predictions for, we can use the batch prediction service.

IBM Watson Developer Cloud

Analyze text to extract meta-data from content such as concepts, entities, keywords, categories, sentiment, emotion, relations, semantic roles, using natural language understanding. With custom annotation models developed using Watson Knowledge Studio, identify industry/domain specific entities and relations in unstructured text.

Programming Language Supported:

Java, python, swift, Node.js, .NET, Unity

Natural language API supports following features:

Concepts - Identify high-level concepts that aren't necessarily directly referenced in the text.

Entities - Find people, places, events, and other types of entities mentioned in your content.

Keywords - Search our content for relevant keywords.

Categories - Categorize your content using a five-level classification hierarchy.

Sentiment

Emotion - Analyze emotion conveyed by specific target phrases or by the document as a whole. We can also enable emotion analysis for entities and keywords that are automatically detected by the service.

Relations – Recognize when two entities are related, and identify the type of relation.

Metadata - For HTML and URL input, get the author of the webpage, the page title, and the publication date.

Semantic roles - Parse sentences into subject-action-object form, and identify entities and keywords that are subjects or objects of an action.

Languages Supported:

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Sentiment | Semantic Roles | Relations | Metadata | Keywords | Entities | Emotion | Concepts | Categories |
| Arabic | X |  | X | X |  |  |  |  | X |
| English | X | X | X | X | X | X | X | X | X |
| French | X |  | X | X | X | X |  |  | X |
| German | X |  | X | X | X | X |  |  |  |
| Italian | X |  | X | X | X | X |  |  | X |
| Japanese |  |  |  | X |  |  |  |  |  |
| Portuguese | X |  | X | X | X | X |  |  | X |
| Russian | X |  |  | X | X | X |  |  |  |
| Spanish | X | X | X | X | X | X |  | X | X |
| Swedish |  |  |  | X | X | X |  |  |  |

**Pricing:**

Price is calculated in terms of items. One item is one feature with up to 10,000 characters with a maximum file size of 50kb. Features include Categories, Concepts, Emotion, Entities, Keywords, Metadata, Relations, Semantic Roles, and Sentiment. Each feature in the call is charged as a separate item. For example, if you specify the service to give you Concepts and Entities on text that the service cleans to 8,000 characters, then that costs two items - for the two features. However, if the cleaned text has 11,000 characters, then this call costs you four items, because the service splits calls with more than 10,000 characters into multiple items.

|  |  |
| --- | --- |
| Free Plan | Standard Plan |
| Analyze upto 1000 items per day | 1 – 250,000 items per month  $0.0003 per item |
| Use against any of the features.  Stores one free custom model. | 250,001 – 5,000,000 items per month  $0.0001 per item  5,000,000+ items per month $0.0002 per item Custom model price per month $800 per model There is no daily limit |

Model Deployment:

Model training and deployment is done in Watson Knowledge Studio. Once we are satisfied with the performance of the model (annotator component), we can deploy a version of it to IBM Watson Knowledge Studio and use the deployed custom model. Once available, we need to make a note of the model ID (model\_id). We will provide this ID to the Natural Language Understanding service to enable the service to use your custom model.

Pre-requisites:

* NL Understanding service advanced plan
* IBM Bluemix space and instance name associated with it

To use the deployed model, you must specify the model ID of your custom model in the entities.model parameter.

You can use the model with the Natural Language Understanding GET /analyze request to extract the various features.

Send API requests to the Analyze endpoint using text, HTML, or a public URL, and specify one or more of the following features.

Input: Text, HTML, URL

Output: JSON object

Sample input:

