Datumbox API Documentation 1.0v

1. Introduction

The Datumbox API is a web service which allows you to use our Machine Learning platform from your website, software or mobile application. The API gives you access to all of the supported functions of our service. The Web Service uses "REST-Like" RPC-style operations over HTTP POST requests with parameters URL encoded into the request and its response is encoded in JSON. It is designed to be easy to use and you can implement it in any model computer language that allows you generating web requests.

In order to use the Datumbox API, you must sign-up for an account, login and get your API Key from your member area and download from our API page this documentation and code samples which will help you build your implementation. Additionally while developing your code it is advised to check out our API Sandbox and generate some test requests.

Using the Datumbox API or the Website indicates that you have read and accept the Terms & Conditions and the Privacy Policy. If you do not accept these terms, you are not authorized to use this service.

2. Building Applications with our API

Our API allows you to build your own intelligent applications within minutes. There are no limitations on what type of applications that you can build (web services, mobile apps, windows apps etc). All you have to do is to call our service when you need to use the Machine Learning functionalities.

Currently our API allows you to build applications that make use of Text Analysis and Natural Language Processing techniques such as Online Marketing Tools, SEO Tools, Social Media Monitoring services, Anti-Spam filters and other Text Classification apps. The currently supported API functions are: Sentiment Analysis, Subjectivity Analysis, Topic Classification, Spam Detection, Adult Content Detection, Readability Assessment, Language Detection, Commercial Detection, Educational Detection, Gender Detection, Keyword Extraction, Text Extraction and Document Similarity.

3. Supported API functions

Currently the Datumbox API is on 1.0 version and supports 14 functions:

3.1 Sentiment Analysis

HTTP Method: POST
URL: http://api.datumbox.com/1.0/SentimentAnalysis.json
Description: The Sentiment Analysis function classifies documents as positive, negative or neutral (lack of sentiment) depending on whether they express a positive, negative or neutral opinion.

Request Parameters:

Parameter Name	Description
api_key	Your API Key
text	The clear text (no HTML tags) that we evaluate.

JSON Response:

```
{
  "output": {
    "status": 1,
    "result": "positive"
  }
}
```

JSON "result" value: In this function the possible values of the above JSON result field are: "positive", "negative" or "neutral".

3.2 Twitter Sentiment Analysis

HTTP Method: POST
URL: http://api.datumbox.com/1.0/TwitterSentimentAnalysis.json
Description: The Twitter Sentiment Analysis function allows you to perform Sentiment Analysis on Twitter. It classifies the tweets as positive, negative or neutral depending on their context.

Request Parameters:

Parameter Name	Description
api_key	Your API Key
text	The text of the tweet that we evaluate.

JSON Response:

```
{
  "output": {
    "status": 1,
    "result": "positive"
  }
}
```

JSON "result" value: In this function the possible values of the above JSON result field are: "positive", "negative" or "neutral".

3.3 Subjectivity Analysis

HTTP Method: POST
URL: http://api.datumbox.com/1.0/SubjectivityAnalysis.json
Description: The Subjectivity Analysis function categorizes documents as subjective or objective based on their writing style. Texts that express personal opinions are labeled as subjective and the others as objective.

Request Parameters:

Parameter Name	Description
api_key	Your API Key
text	The clear text (no HTML tags) that we evaluate.

JSON Response:

```
{
  "output": {
    "status": 1,
    "result": "objective"
  }
}
```

JSON "result" value: In this function the possible values of the above JSON result field are: "objective" or "subjective".

3.4 Topic Classification

HTTP Method: POST
URL: http://api.datumbox.com/1.0/TopicClassification.json
Description: The Topic Classification function assigns documents in 12 thematic categories based on their keywords, idioms and jargon. It can be used to identify the topic of the texts.

Request Parameters:

Parameter Name	Description
api_key	Your API Key
text	The clear text (no HTML tags) that we evaluate.

JSON Response:

```
{
  "output": {
    "status": 1,
    "result": " Science"
  }
}
```

JSON "result" value: In this function the possible values of the above JSON result field are: "Arts", "Business & Economy", "Computers & Technology", "Health", "Home & Domestic Life", "News", "Recreation & Activities", "Reference & Education", "Science", "Shopping", "Society" or "Sports".

3.5 Spam Detection

HTTP Method: POST
URL: http://api.datumbox.com/1.0/SpamDetection.json
Description: The Spam Detection function labels documents as spam or nospam by taking into account their context. It can be used to filter out spam emails and comments.

Request Parameters:

Parameter Name	Description
api_key	Your API Key
text	The clear text (no HTML tags) that we evaluate.

JSON Response:

```
{
  "output": {
    "status": 1,
    "result": "nospam"
  }
}
```

JSON "result" value: In this function the possible values of the above JSON result field are: "spam" or "nospam".

3.6 Adult Content Detection

HTTP Method: POST
URL: http://api.datumbox.com/1.0/AdultContentDetection.json
Description: The Adult Content Detection function classifies the documents as adult or noadult based on their context. It can be used to detect whether a document contains content unsuitable for minors.

Request Parameters:

Parameter Name	Description
api_key	Your API Key
text	The clear text (no HTML tags) that we evaluate.

JSON Response:

```
{
  "output": {
    "status": 1,
    "result": "noadult"
  }
}
```

JSON "result" value: In this function the possible values of the above JSON result field are: "adult" or "noadult".

3.7 Readability Assessment

HTTP Method: POST
URL: http://api.datumbox.com/1.0/ReadabilityAssessment.json
Description: The Readability Assessment function determines the degree of readability of a document based on its terms and idioms. The texts are classified as basic, intermediate and advanced depending their difficulty.

Request Parameters:

Parameter Name	Description
api_key	Your API Key
text	The clear text (no HTML tags) that we evaluate.

JSON Response:

```
{
  "output": {
    "status": 1,
    "result": "basic"
  }
}
```

JSON "result" value: In this function the possible values of the above JSON result field are: "basic", "intermediate" or "advanced".

3.8 Language Detection

HTTP Method: POST
URL: http://api.datumbox.com/1.0/LanguageDetection.json
Description: The Language Detection function identifies the natural language of the given document based on its words and context. This classifier is able to detect 96 different languages.

Request Parameters:

Parameter Name	Description
api_key	Your API Key
text	The clear text (no HTML tags) that we evaluate.

JSON Response:

```
{
  "output": {
    "status": 1,
    "result": "en"
  }
}
```

JSON "result" value: In this function the possible values of the JSON result field are the ISO639-1 two-letter language codes. The exhaustive list of possible value is the following: "af", "ar", "az", "bcl", "be-x-old", "bg", "bn", "br", "ca", "ceb", "cs", "cy", "da", "de", "diq", "dv", "el", "en", "es", "et", "eu", "fa", "fi", "fr", "frr", "fur", "fy", "ga", "gl", "gn", "gu", "gv", "he", "hi", "hif", "hr", "hsb", "ht", "hu", "hy", "id", "ilo", "io", "is", "it", "jbo", "ka", "kn", "ko", "ksh", "ku", "kw", "ln", "lt", "lv", "mg", "mi", "ml", "mr", "ms", "mt", "my", "nah", "nl", "nn", "no", "oc", "or", "pam", "pl", "pms", "pnb", "pt", "ro", "ru", "se", "sk", "sl", "so", "sq", "sr", "stq", "sv", "sw", "ta", "te", "tk", "tl", "tr", "uk", "ur", "uz", "vi", "wa", "war" and "yo".

3.9 Commercial Detection

HTTP Method: POST
URL: http://api.datumbox.com/1.0/CommercialDetection.json
Description: The Commercial Detection function labels the documents as commercial or non-commercial based on their keywords and expressions. It can be used to detect whether a website is commercial or not.

Request Parameters:

Parameter Name	Description
api_key	Your API Key
text	The clear text (no HTML tags) that we evaluate.

JSON Response:

```
{
  "output": {
    "status": 1,
    "result": "noncommercial"
  }
}
```

JSON "result" value: In this function the possible values of the above JSON result field are: "commercial" or "noncommercial".

3.10 Educational Detection

HTTP Method: POST
URL: http://api.datumbox.com/1.0/EducationalDetection.json
Description: The Educational Detection function classifies the documents as educational or non-educational based on their context. It can be used to detect whether a website is educational or not.

Request Parameters:

Parameter Name	Description
api_key	Your API Key
text	The clear text (no HTML tags) that we evaluate.

JSON Response:

```
{
  "output": {
    "status": 1,
    "result": "noneducational"
  }
}
```

JSON "result" value: In this function the possible values of the above JSON result field are: "educational" or "noneducational".

3.11 Gender Detection

HTTP Method: POST
URL: http://api.datumbox.com/1.0/GenderDetection.json
Description: The Gender Detection function identifies if a particular document is written-by or targets-to a man or a woman based on the context, the words and the idioms found in the text.

Request Parameters:

Parameter Name	Description
api_key	Your API Key
text	The clear text (no HTML tags) that we evaluate.

JSON Response:

```
{
  "output": {
    "status": 1,
    "result": "male"
  }
}
```

JSON "result" value: In this function the possible values of the above JSON result field are: "male" or "female".

3.12 Keyword Extraction

HTTP Method: POST
URL: http://api.datumbox.com/1.0/KeywordExtraction.json
Description: The Keyword Extraction function enables you to extract from an arbitrary document all the keywords and word-combinations along with their occurrences in the text.

Request Parameters:

Parameter Name	Description
api_key	Your API Key
n	The number of Keyword Combinations (n-grams) that we want to extract.
text	The clear text (no HTML tags) that we evaluate.

JSON Response:

```
{
  "output": {
    "status": 1,
    "result": {
      "1": {
        "this": 1,
        "is": 1,
        "an": 1,
        "example": 1,
        "text": 1
      },
      "2": {
        "this is": 1,
        "is an": 1,
        "an example": 1,
        "example text": 1
      }
    }
  }
}
```

JSON "result" value: In this function the result stores the array of keyword lists, having one list for every keyword combination. The Keyword Lists use key/value format where in key we store the keyword combination and on value the number of its occurrences within the text.

3.13 Text Extraction

HTTP Method: POST
URL: http://api.datumbox.com/1.0/TextExtraction.json
Description: The Text Extraction function enables you to extract the important information from a given webpage. Extracting the clear text of the documents is an important step before any other analysis.

Request Parameters:

Parameter Name	Description
api_key	Your API Key
text	The HTML source of the webpage.

JSON Response:

```
{
  "output": {
    "status": 1,
    "result": "this is the clear text of the html page"
  }
}
```

JSON "result" value: In this function the result stores the clear text of the webpage.

3.14 Document Similarity

HTTP Method: POST
URL: http://api.datumbox.com/1.0/DocumentSimilarity.json
Description: The Document Similarity function estimates the degree of similarity between two documents. It can be used to detect duplicate webpages or detect plagiarism.

Request Parameters:

Parameter Name	Description
api_key	Your API Key
original	The first clear text (no HTML tags) that we compare
copy	The second clear text (no HTML tags) that we compare

JSON Response:

```
{
  "output": {
    "status": 1,
    "result": {
      "Oliver": 0.78260869565217,
      "Shingle": 0.2
    }
  }
}
```

JSON "result" value: In this function the result stores two different document similarity metrics which store values from 0 (no similarity) to 1 (exactly equal).

4. Error Messages

If an error occurs while executing your call, our API will return a status of 0. In that case it results also the error code and the error message. Here is how an Error Reply looks like:

```
{
  "output": {
    "status": 0,
    "error": {
      "ErrorCode": 6,
      "ErrorMessage": "Invalid Account"
    }
  }
}
```

Here is the complete list of Error Codes and Error Messages:

Error Code	Error Message
1	Invalid API Version
2	Invalid API Function
3	Missing Parameter
4	Unexpected Error
5	The timestamp expired
6	Invalid Account
7	Your account is not Active
8	Invalid Authentication Key
9	Invalid Subscription
10	You can't access the particular Service. Please use a different Subscription.
11	You reached your daily limit: X API calls

5. Rate Limits

In Datumbox the rate limits are imposed as a limit on the number of API calls that can be generated during a specific time period (window of time). The most common rate limit that is used in Datumbox is the maximum number of daily API calls; nevertheless the exact time period and limitations depend on the Subscription of the user. The basic free subscription allows you to generate 1000 calls per day.

6. Useful Resources

Before you start building your API client make sure you visit our API page to get the newest version of this documentation and download our code samples. Moreover before developing your API client, have a look on our API Sandbox which will allow you to test our API without needing to write any code.

Please use the following Contact Form to report any bugs.