文本分析

Text Analytics API是一套使用Azure机器学习构建的文本分析Web服务。该API 可以用于分析非结构化的文本，完成情感分析、 关键词提取和语言检测等任务。使用该API不需要训练数据; 只需你带来的文本数据。 该API使用高级自然语言处理技术来提供最佳的类预测。

* 语言检测（Detect Language）

1. Request URL

[https://westus.api.cognitive.microsoft.com/text/analytics/v2.0/languages[?numberOfLanguagesToDetect](https://westus.api.cognitive.microsoft.com/text/analytics/v2.0/languages%5b?numberOfLanguagesToDetect)

1. Request parameters
2. numberOfLanguagesToDetect (integer，可选)语言检测数目默认为1.
3. Request body

{

"type": "object",

"properties": {

"documents": {

"type": "array",

"items": {

"type": "object",

"properties": {

"id": {

"description": "Unique, non-empty document identifier.",

"type": "string"

},

"text": {

"type": "string"

}

}

}

}

}

}

1. 返回值（JSON）

{

"type": "object",

"properties": {

"documents": {

"type": "array",

"items": {

"type": "object",

"properties": {

"id": {

"description": "Unique document identifier.",

"type": "string"

},

"detectedLanguages": {

"description": "A list of extracted languages.",

"type": "array",

"items": {

"type": "object",

"properties": {

"name": {

"description": "Long name of a detected language (e.g. English, French).",

"type": "string"

},

"iso6391Name": {

"description": "A two letter representation of the detected language according to the ISO 639-1 standard (e.g. en, fr).",

"type": "string"

},

"score": {

"format": "double",

"description": "A confidence score between 0 and 1. Scores close to 1 indicate 100% certainty that the identified language is true.",

"type": "number"

}

}

}

}

}

}

},

"errors": {

"type": "array",

"items": {

"type": "object",

"properties": {

"id": {

"description": "Input document unique identifier the error refers to.",

"type": "string"

},

"message": {

"description": "Error message.",

"type": "string"

}

}

}

}

}

}

* 主题检测（Detect Topics）

API会返回提交的文本文档列表的检测到的主题。 使用关键短语来标识主题，关键短语可以是一个或多个相关词。 使用网址参数和停用词列表控制过滤掉哪些字或文档。 您还可以提供要从响应中排除的主题列表。 必须提交至少100个文本文档，但是它设计为检测跨越数百到数千个文档的主题。 请注意，每个提交的文本文档收取一笔交易费用。 为了获得最佳性能，请将每个文档限制为一个简短的人性书面文本段落，例如评论，对话或用户反馈。

1. Request URL

[https://westus.api.cognitive.microsoft.com/text/analytics/v2.0/topics[?minDocumentsPerWord][&maxDocumentsPerWord](https://westus.api.cognitive.microsoft.com/text/analytics/v2.0/topics%5b?minDocumentsPerWord%5d%5b&maxDocumentsPerWord)]

1. Request parameters
   1. minDocumentsPerWord(integer,可选)使用此参数可帮助排除罕见文档主题。选择适当的值来进行忽略服务。
   2. maxDocumentsPerWord(integer,可选) 使用此参数可帮助排除无处不在的文档主题。 选择适当的值来进行忽略服务。
2. Request body

{

"type": "object",

"properties": {

"stopWords": {

"description": "List of words to ignore from all documents during pre-processing.",

"type": "array",

"items": {

"type": "string"

}

},

"topicsToExclude": {

"description": "List of topics to omit from the response.",

"type": "array",

"items": {

"type": "string"

}

},

"documents": {

"type": "array",

"items": {

"type": "object",

"properties": {

"id": {

"description": "Unique, non-empty document identifier.",

"type": "string"

},

"text": {

"type": "string"

}

}

}

}

}

}

* 关键词（Key Phrases）

API返回一个字符串列表，表示输入文本中的关键会话点。 我们使用来自Microsoft Office的复杂自然语言处理工具包的技术。 目前，支持以下语言：英语，德语，西班牙语和日语。

1. Request URL

<https://westus.api.cognitive.microsoft.com/text/analytics/v2.0/keyPhrases>

1. Request body

{

"type": "object",

"properties": {

"documents": {

"type": "array",

"items": {

"type": "object",

"properties": {

"language": {

"description": "This is the 2 letter ISO 639-1 representation of a language.\r\n For example, use \"en\" for English; \"es\" for Spanish etc.,",

"type": "string"

},

"id": {

"description": "Unique, non-empty document identifier.",

"type": "string"

},

"text": {

"type": "string"

}

}

}

}

}

}

1. 返回值（JSON）

{

"type": "object",

"properties": {

"documents": {

"type": "array",

"items": {

"type": "object",

"properties": {

"keyPhrases": {

"description": "A list of representative words or phrases. The number of key phrases returned is proportional to the number of words in the input document.",

"type": "array",

"items": {

"type": "string"

}

},

"id": {

"description": "Unique document identifier.",

"type": "string"

}

}

}

},

"errors": {

"type": "array",

"items": {

"type": "object",

"properties": {

"id": {

"description": "Input document unique identifier the error refers to.",

"type": "string"

},

"message": {

"description": "Error message.",

"type": "string"

}

}

}

}

}

}

* 操作状态（Operation Status）

获取提交进行处理的操作的状态。 如果操作已达到“Succeeded”状态，也将返回结果。

1. Request URL

[https://westus.api.cognitive.microsoft.com/text/analytics/v2.0/operations/{operationId}](https://westus.api.cognitive.microsoft.com/text/analytics/v2.0/operations/%7boperationId%7d)

1. Request parameters
2. operationId (string) 已提交操作的唯一ID。
3. 返回值

{

"type": "object",

"properties": {

"status": {

"description": "Operation status.",

"enum": [

"notStarted",

"running",

"failed",

"cancelled",

"succeeded"

],

"type": "string"

},

"createdDateTime": {

"format": "date-time",

"description": "Operation creation date time (ISO 8601 literal).",

"type": "string"

},

"lastActionDateTime": {

"format": "date-time",

"description": "Operation last status change date time (ISO 8601 literal).",

"type": "string"

},

"operationType": {

"description": "Name of API endpoint that created the operation.",

"type": "string"

},

"operationProcessingResult": {

"required": [

"discriminator"

],

"type": "object",

"properties": {

"errors": {

"type": "array",

"items": {

"type": "object",

"properties": {

"id": {

"description": "Input document unique identifier the error refers to.",

"type": "string"

},

"message": {

"description": "Error message.",

"type": "string"

}

}

}

},

"discriminator": {

"type": "string"

}

},

"discriminator": "discriminator",

"description": "Operation result. Specific format varies according to the operation type. Exists only in case the operation has reached a 'Succeeded' state."

},

"message": {

"description": "Error message. Exists only in case the operation has reached a 'Failed' state.",

"type": "string"

}

}

}

* 情绪（Sentiment）

API返回0和1之间的数字分数。接近1的分数表示积极情绪，接近0的分数表示消极情绪。 情感分数是使用分类技术生成的。 分类器的输入特征包括n-gram，从词性标签生成的特征和词嵌入。 目前，支持以下语言：英语，西班牙语，法语，葡萄牙语。

1. Request URL

<https://westus.api.cognitive.microsoft.com/text/analytics/v2.0/sentiment>

1. Request body

{

"type": "object",

"properties": {

"documents": {

"type": "array",

"items": {

"type": "object",

"properties": {

"language": {

"description": "This is the 2 letter ISO 639-1 representation of a language.\r\n For example, use \"en\" for English; \"es\" for Spanish etc.,",

"type": "string"

},

"id": {

"description": "Unique, non-empty document identifier.",

"type": "string"

},

"text": {

"type": "string"

}

}

}

}

}

}

3．返回值（JSON）

{

"type": "object",

"properties": {

"documents": {

"type": "array",

"items": {

"type": "object",

"properties": {

"score": {

"format": "double",

"description": "A decimal number between 0 and 1 denoting the sentiment of the document. \r\n A score above 0.7 usually refers to a positive document while a score below 0.3 normally has a negative connotation.\r\n Mid values refer to neutral text.",

"type": "number"

},

"id": {

"description": "Unique document identifier.",

"type": "string"

}

}

}

},

"errors": {

"type": "array",

"items": {

"type": "object",

"properties": {

"id": {

"description": "Input document unique identifier the error refers to.",

"type": "string"

},

"message": {

"description": "Error message.",

"type": "string"

}

}

}

}

}

}