

Desarrollo de soluciones de procesamiento del lenguaje natural con Servicios de Azure AI

Las soluciones de procesamiento de lenguaje natural (NLP) usan modelos de lenguaje para interpretar el significado semántico del lenguaje escrito o hablado. Puede usar el servicio Language Understanding a fin de compilar modelos de lenguaje para las aplicaciones.

Requisitos previos

Antes de iniciar esta ruta de aprendizaje, debe contar con lo siguiente:

Familiaridad con Azure y Azure Portal

Experiencia en programación con C# en Python. Si no tiene ninguna experiencia anterior en programación, se recomienda completar la ruta de aprendizaje Primeros pasos con C# o Primeros pasos con Python antes de empezar con esta.

Introducción

Cada día, en todo el mundo se genera una cantidad enorme de datos, gran parte de la cual es texto en forma de correos electrónicos, publicaciones en redes sociales, reseñas en línea, documentos empresariales y mucho más. Las técnicas de inteligencia artificial que aplican modelos estadísticos y semánticos permiten crear aplicaciones que extraen información y significado de estos datos basados en texto.

Lenguaje de Azure AI proporciona una API para las técnicas comunes de análisis de texto que puede integrar fácilmente en el código de su propia aplicación

En este módulo, aprenderá a usar el servicio Lenguaje de Azure AI para

Detectar el idioma del texto.

Analizar opiniones de texto.

Extraer frases clave, entidades y entidades vinculadas.

Aprovisionar un recurso de Lenguaje de Azure AI

Lenguaje de Azure AI está diseñado para ayudarle a extraer información del texto. Proporciona funcionalidad que puede usar para:

Detección de idioma: determina el idioma en el que está escrito el texto.

Extracción de frases clave: determina en el texto las frases y palabras importantes que indican los puntos principales.

Análisis de sentimiento: cuantifica qué tan positivo o negativo que es el texto.

Reconocimiento de entidades con nombre: detecta las referencias a entidades, incluidas personas, ubicaciones, períodos de tiempo, organizaciones y más.

Vinculación de entidad: identifica entidades específicas al proporcionar vínculos de referencia a artículos de Wikipedia.

Recursos de Azure para el análisis de texto

Para usar Lenguaje de Azure AI, debe aprovisionar un recurso para él en su suscripción de Azure.

Después de aprovisionar un recurso adecuado en la suscripción de Azure, puede usar su punto de conexión y una de sus claves para llamar a las API de Lenguaje de Azure AI desde el código. Puede llamar a las API de Lenguaje de Azure AI mediante el envío de solicitudes en formato JSON a la interfaz REST o por medio de cualquiera de los SDK específicos del lenguaje de programación disponibles.

Detectar idioma

La API Detección de idioma de Azure AI evalúa la entrada de texto y, para cada documento enviado, devuelve identificadores de idioma con una puntuación que indica la solidez del análisis.

Esta capacidad es útil para los almacenes de contenido que recopilan texto arbitrario, donde el idioma es desconocido. Otro escenario podría incluir un bot de chat. Si un usuario inicia una sesión con el bot de chat, la detección de idioma se puede usar para determinar el idioma que usa, y permite configurar las respuestas del bot en el idioma adecuado.

Puede analizar los resultados del análisis para determinar el idioma que se usa en el documento de entrada. La respuesta también devuelve una puntuación que refleja la confianza del modelo (un valor entre 0 y 1).

La detección de idioma puede funcionar con documentos o frases únicas. Es importante tener en cuenta que el documento debe tener un tamaño inferior a 5120 caracteres. El límite de tamaño es por documento, y cada colección está restringida a 1000 elementos (identificadores). Aquí se muestra un ejemplo de una carga JSON con el formato adecuado que podría enviar al servicio en el cuerpo de la solicitud, incluida una colección de documentos, donde cada uno de los cuales contiene un identificador único y el texto que se va a analizar. También puede proporcionar una propiedad `countryHint` para mejorar el rendimiento de la predicción.

```
{"kind":"LangDetect","params":{"modelVer":"latest"},"input":{"docs":[{"id":"1","text":"Hello world","hint":"US"}, {"id":"2","text":"Bonjour tout le monde"}]}}
```

El servicio devolverá una respuesta JSON que contiene un resultado para cada documento en el cuerpo de la solicitud, incluido el idioma previsto y un valor que indica el nivel de confianza de la predicción. El nivel de confianza es un valor comprendido entre 0 y 1, donde los valores más cercanos a 1 tienen un nivel de confianza mayor. A continuación, se muestra un ejemplo de una respuesta JSON estándar que se asigna al JSON de la solicitud anterior.

```
{"kind":"LangDetectResults","results":{"docs":[{"lang":{"score":1,"iso":"en","name":"English"},"id":"1","warn":[]}, {"lang":{"score":1,"iso":"fr","name":"French"},"id":"2","warn":[]}],"errors":[],"modelVer":"2022-10-01"}}
```

En nuestro ejemplo, todos los idiomas muestran una confianza de 1, principalmente porque el texto es relativamente sencillo y su idioma es fácil de detectar.

Si pasa un documento que tiene contenido en varios idiomas, el servicio tendrá un comportamiento algo distinto. El contenido en varios idiomas dentro del mismo documento devuelve el idioma con mayor representación en el contenido, pero con una clasificación positiva inferior, que refleja la solidez marginal de esa evaluación. En el ejemplo siguiente, la

entrada es una mezcla de inglés, español y francés. El analizador usa el análisis estadístico del texto para determinar el idioma predominante.

```
{"docs":[{"id":"1","text":"Hello, I would like to take a class at your University. ¿Se ofrecen clases en español? Es mi primera lengua y más fácil para escribir. Que diriez-vous des cours en français?"}]}
```

En el ejemplo siguiente se muestra una respuesta para este ejemplo de varios idiomas.

```
{"docs":[{"id":"1","lang":{"name":"Spanish","iso":"es","score":0.9375},"warn":[]},"errors":[],"modelVer":"2022-10-01"}
```

La última condición que se debe tener en cuenta es cuando existe ambigüedad en cuanto al idioma del contenido. Este escenario puede producirse si envía contenido textual que el analizador no puede analizar; por ejemplo, debido a problemas de codificación de caracteres al convertir el texto en una variable de cadena. Como resultado, la respuesta para el nombre y el código ISO del idioma indicará "(Desconocido)" y el valor de puntuación se devolverá como 0. En el ejemplo siguiente se muestra el aspecto de la respuesta.

```
{"docs":[{"id":"1","lang":{"name":"(Unknown)","iso":"(Unknown)","score":0.0},"warn":[]},"errors":[],"modelVer":"2022-10-01"}
```

Extracción de frases clave

La extracción de frases clave es el proceso de evaluar el texto de uno o varios documentos e identificar los puntos principales del contexto de dichos documentos.

La extracción de frases clave funciona mejor con documentos más grandes (el tamaño máximo que se puede analizar es de 5120 caracteres).

Al igual que lo que ocurre con la detección de idioma, la interfaz REST permite enviar uno o varios documentos para su análisis.

```
{"kind":"KeyPhraseExtraction","params":{"modelVer":"latest"},"input":{"docs":[{"id":"1","lang":"en","text":"You must be the change you wish to see in the world."},{id":"2","lang":"en","text":"The journey of a thousand miles begins with a single step."}]}}
```

La respuesta contiene una lista de frases clave detectadas en cada documento:

```
{"kind":"KeyPhraseExtractionResults","results":{"docs":[{"id":"1","keyPhrases":["change","world"],"warn":[]},{id":"2","keyPhrases":["miles","single step","journey"],"warn":[]},"errors":[],"modelVer":"2021-06-01"}}
```

Análisis de opinión

El análisis de sentimiento se usa para evaluar qué tan positivo o negativo es un documento de texto, lo que puede resultar útil en varias cargas de trabajo, como:

Al evaluar una película, libro o producto mediante la cuantificación de las opiniones basadas en reseñas.

Al priorizar las respuestas del servicio de atención al cliente a la correspondencia recibida a través de correo electrónico o los mensajes de redes sociales.

Cuando se usa Lenguaje de Azure AI para evaluar las opiniones, la respuesta incluye la opinión general del documento y la opinión de oraciones individuales de cada documento enviado al servicio.

Por ejemplo, podría enviar un único documento para el análisis de sentimiento de la manera siguiente:

```
{"kind":"SentimentAnalysis","params":{"modelVer":"latest"},"input":{"docs":[{"id":"1","lang":"en","text":"Good morning!"}]}}
```

La respuesta del servicio podría ser similar a esta:

```
{"kind":"SentimentAnalysisResults","results":{"docs":[{"id":"1","sentiment":"positive","scores":{"positive":0.89,"neutral":0.1,"negative":0.01},"sentences":[{"sentiment":"positive","scores":{"positive":0.89,"neutral":0.1,"negative":0.01},"offset":0,"length":13,"text":"Good morning!"}],"warn":[]},"errors":[],"modelVer":"2022-11-01"}}
```

La opinión de la oración se basa en las puntuaciones de confianza para los valores de clasificación positivos, negativos y neutros entre 0 y 1.

La opinión general del documento se basa en oraciones:

Si todas las oraciones son neutras, la opinión general es neutra.

Si las clasificaciones de oraciones incluyen solo positivas y neutras, la opinión general es positiva.

Si las clasificaciones de oraciones incluyen solo negativas y neutras, la opinión general es negativa.

Si las clasificaciones de oraciones incluyen las positivas y las negativas, la opinión general es mixta.

Extraer entidades

El reconocimiento de entidades con nombre identifica las entidades que se mencionan en el texto. Las entidades se agrupan en categorías y subcategorías, por ejemplo:

Person

Location

DateTime

Organización

Dirección

Correo electrónico

URL

La entrada para el reconocimiento de entidades es similar a la entrada para otras funciones de API de Lenguaje de Azure AI:

```
{"kind":"EntityRecognition","params":{"modelVer":"latest"},"input":{"docs":[{"id":"1","lang":"en","text":"Joe went to London on Saturday"}]}}
```

La respuesta incluye una lista de entidades categorizadas que se encuentran en cada documento:

```
{"kind":"EntityRecognitionResults","results":{"docs":[{"entities":[{"text":"Joe","category":"Person","offset":0,"length":3,"score":0.62}, {"text":"London","category":"Location","sub":"GPE","offset":12,"length":6,"score":0.88}, {"text":"Saturday","category":"DateTime","sub":"Date","offset":22,"length":8,"score":0.8}], "id":"1", "warn":[]}, {"errors":[],"modelVer":"2021-01-15"}]}
```

Para obtener más información sobre las entidades, consulte el módulo Crear un modelo de reconocimiento del lenguaje conversacional.

Extracción de entidades vinculadas

En algunos casos, se podría aplicar el mismo nombre a más de una entidad. Por ejemplo, ¿una instancia de la palabra "Venus" hace referencia al planeta o a la diosa de la mitología?

La vinculación de entidad se puede usar para eliminar la ambigüedad respecto de entidades con el mismo nombre mediante una referencia a un artículo de una knowledge base. Wikipedia proporciona la knowledge base para el servicio Text Analytics. Los vínculos a artículos específicos se determinan en función del contexto de la entidad dentro del texto.

Por ejemplo, "Vi a Venus brillando en el cielo" se asocia con el vínculo <https://en.wikipedia.org/wiki/Venus>, mientras que "Venus, la diosa de la belleza" se asocia con [https://en.wikipedia.org/wiki/Venus_\(mythology\)](https://en.wikipedia.org/wiki/Venus_(mythology)).

Al igual que ocurre con todas las funciones del servicio Lenguaje de Azure AI, puede enviar uno o varios documentos para su análisis:

```
{"kind":"EntityLinking","params":{"modelVer":"latest"},"input":{"docs":[{"id":"1","lang":"en","text":"I saw Venus shining in the sky"}]}}
```

La respuesta incluye las entidades identificadas en el texto junto con vínculos a artículos asociados:

```
{"kind":"EntityLinkingResults","results":{"docs":[{"id":"1","entities":[{"bingId":"89253af3-5b63-e620-9227-f839138139f6","name":"Venus","matches":[{"text":"Venus","offset":6,"length":5,"score":0.01}], "lang":"en","id":"Venus","url":"https://en.wikipedia.org/wiki/Venus","dataSource":"Wikipedia"}], "warn":[]}, {"errors":[],"modelVer":"2021-06-01"}]}
```

Resumen

En este módulo, aprendió a utilizar Lenguaje de Azure AI para:

Detectar el idioma del texto.

Analizar opiniones de texto.

Extraer frases clave, entidades y entidades vinculadas.

Para obtener más información sobre el Lenguaje de Azure AI y algunos de los conceptos descritos en este módulo, puede explorar lo siguiente:

Documentación del lenguaje de Azure AI

Creación de un modelo de reconocimiento del lenguaje conversacional

Crear una solución de extracción de entidades con nombre personalizada