

Get started with text analysis


10 minutes

Azure AI Language is a part of the Azure AI services offerings that can perform advanced natural language processing over unstructured text. Azure AI Language's text analysis features include:

- **Named entity recognition** identifies people, places, events, and more. This feature can also be customized to extract custom categories.
- **Entity linking** identifies known entities together with a link to Wikipedia.
- **Personal identifying information (PII) detection** identifies personally sensitive information, including personal health information (PHI).
- **Language detection** identifies the language of the text and returns a language code such as "en" for English.
- **Sentiment analysis and opinion mining** identifies whether text is positive or negative.
- **Summarization** summarizes text by identifying the most important information.
- **Key phrase extraction** lists the main concepts from unstructured text.

Entity recognition and linking

You can provide Azure AI Language with unstructured text and it will return a list of *entities* in the text that it recognizes. An entity is an item of a particular type or a category; and in some cases, subtype, such as those as shown in the following table.

 Expand table

Type	SubType	Example
Person		"Bill Gates", "John"
Location		"Paris", "New York"
Organization		"Microsoft"

Type	SubType	Example
Quantity	Number	"6" or "six"
Quantity	Percentage	"25%" or "fifty percent"
Quantity	Ordinal	"1st" or "first"
Quantity	Age	"90 day old" or "30 years old"
Quantity	Currency	"10.99"
Quantity	Dimension	"10 miles", "40 cm"
Quantity	Temperature	"45 degrees"
DateTime		"6:30PM February 4, 2012"
DateTime	Date	"May 2nd, 2017" or "05/02/2017"
DateTime	Time	"8am" or "8:00"
DateTime	DateRange	"May 2nd to May 5th"
DateTime	TimeRange	"6pm to 7pm"
DateTime	Duration	"1 minute and 45 seconds"
DateTime	Set	"every Tuesday"
URL		" <code>https://www.bing.com</code> "
Email		" <code>support@microsoft.com</code> "
US-based Phone Number		"(312) 555-0176"

Type	SubType	Example
IP Address		"10.0.1.125"

Azure AI Language also supports *entity linking* to help disambiguate entities by linking to a specific reference. For recognized entities, the service returns a URL for a relevant *Wikipedia* article.

For example, suppose you use Azure AI Language to detect entities in the following restaurant review extract:

"I ate at the restaurant in Seattle last week."

[Expand table](#)

Entity	Type	SubType	Wikipedia URL
Seattle	Location		https://en.wikipedia.org/wiki/Seattle
last week	DateTime	DateRange	

Language detection

Use the language detection capability of Azure AI Language to identify the language in which text is written. You can submit multiple documents at a time for analysis. For each document submitted the service will detect:

- The language name (for example "English").
- The ISO 639-1 language code (for example, "en").
- A score indicating a level of confidence in the language detection.

For example, consider a scenario where you own and operate a restaurant where customers can complete surveys and provide feedback on the food, the service, staff, and so on. Suppose you have received the following reviews from customers:

Review 1: *"A fantastic place for lunch. The soup was delicious."*

Review 2: "*Comida maravillosa y gran servicio.*"

Review 3: "*The croque monsieur avec frites was terrific. Bon appetit!*"

You can use the text analytics capabilities in Azure AI Language to detect the language for each of these reviews; and it might respond with the following results:

 Expand table

Document	Language Name	ISO 6391 Code	Score
Review 1	English	en	1.0
Review 2	Spanish	es	1.0
Review 3	English	en	0.9

Notice that the language detected for review 3 is English, despite the text containing a mix of English and French. The language detection service will focus on the ***predominant*** language in the text. The service uses an algorithm to determine the predominant language, such as length of phrases or total amount of text for the language compared to other languages in the text. The predominant language will be the value returned, along with the language code. The confidence score might be less than 1 as a result of the mixed language text.

There might be text that is ambiguous in nature, or that has mixed language content. These situations can present a challenge. An ambiguous content example would be a case where the document contains limited text, or only punctuation. For example, using Azure AI Language to analyze the text ":-)", results in a value of **unknown** for the language name and the language identifier, and a score of **NaN** (which is used to indicate *not a number*).

Sentiment analysis and opinion mining

The text analytics capabilities in Azure AI Language can evaluate text and return sentiment scores and labels for each sentence. This capability is useful for detecting positive and negative sentiment in social media, customer reviews, discussion forums and more.

Azure AI Language uses a prebuilt machine learning classification model to evaluate the text. The service returns sentiment scores in three categories: positive, neutral, and negative. In each of the categories, a score between 0 and 1 is provided. Scores indicate how likely the provided text is a particular sentiment. One document sentiment is also provided.

For example, the following two restaurant reviews could be analyzed for sentiment:

Review 1: "We had dinner at this restaurant last night and the first thing I noticed was how courteous the staff was. We were greeted in a friendly manner and taken to our table right away. The table was clean, the chairs were comfortable, and the food was amazing."

and

Review 2: "Our dining experience at this restaurant was one of the worst I've ever had. The service was slow, and the food was awful. I'll never eat at this establishment again."

The sentiment score for the first review might be: Document sentiment: positive Positive score: .90
Neutral score: .10 Negative score: .00

The second review might return a response: Document sentiment: negative Positive score: .00
Neutral score: .00 Negative score: .99

Key phrase extraction

Key phrase extraction identifies the main points from text. Consider the restaurant scenario discussed previously. If you have a large number of surveys, it can take a long time to read through the reviews. Instead, you can use the key phrase extraction capabilities of the Language service to summarize the main points.

You might receive a review such as:

"We had dinner here for a birthday celebration and had a fantastic experience. We were greeted by a friendly hostess and taken to our table right away. The ambiance was relaxed, the food was amazing, and service was terrific. If you like great food and attentive service, you should try this place."

Key phrase extraction can provide some context to this review by extracting the following phrases:

- birthday celebration
- fantastic experience
- friendly hostess
- great food
- attentive service
- dinner
- table
- ambiance
- place

As well as using sentiment analysis to determine that this is a positive review, you can also use the key phrase service to identify important elements of the review.

Create a resource for Azure AI Language

To use Azure AI Language in an application, you must provision an appropriate resource in your Azure subscription. You can choose either of the following types of resource:

- A **Language** resource - choose this resource type if you only plan to use Azure AI Language services, or if you want to manage access and billing for the resource separately from other services.
- An **Azure AI services** resource - choose this resource type if you plan to use Azure AI Language in combination with other Azure AI services, and you want to manage access and billing for these services together.

Next unit: Exercise - Analyze text with Language Studio

[Continue >](#)