IBM Watson NLU
Natural Language Understanding

Train custom category models to create classifications relevant to your domain

Features

- 5 level category hierarchy
- Train by keyphrases
- JSON input
- Wikipedia based taxonomy
- Explainability

Custom Categories (Beta)

Benefits

Flexible
Make your hierarchy shallow or deep as needed, up to 5 category levels.

Painless training
Train your model in NLU with short key phrases, no labeled data needed.

Explainability
Use explainability to understand why your model predicted a category.

How it Works

Training custom categories

To train a custom categories model, you will need to provide labels and their corresponding “key phrases.” Custom categories then creates a mathematical representation for each label using the key phrases. These mathematical representations are based on mappings to Wikipedia terms.

Analyzing text with your custom categories

At predict time, the categories algorithm computes a mathematical representation for your incoming text and returns the custom categories that are most mathematically similar to it.

Categories also supports hierarchy natively, i.e you can specify labels in a tree fashion like, ‘movies/action.’ In this case, categories will traverse the hierarchy to decide the best possible category node to return.

IBM **Watson**®

IBM Watson NLU
Natural Language Understanding

Custom Categories (Beta)

Details

- API Version: 2021-02-16
- Language Support: English
- Region: US-South

Best Practices

Categories have a distinct mathematical representation that is based on the key phrases used to define them. It's important to ensure key phrases are distinct to avoid confusing results.

Tips for training your model

1. Don't use the same key phrases across different category hierarchies:
   ✕ 'apple' should not be a key phrase for both **/fruit** & **/laptop**

2. Don't overlap words in key phrases across different categories:
   ✕ 'sales' and 'salesman' should not be in different categories

3. Don't use key phrases with similar meanings across different categories:
   ✕ 'cold' and 'chilly' should not be in different categories

4. Don't repeat category labels:
   ✕ **/transportation/car** and **/expenses/car** should not both have 'car' as a category label

Note: Different categories refers to categories with divergent roots. **/fruit** and **/car** are divergent, whereas **/fruit** and **/food/fruit** are within the same category.

IBM **Watson**®

Custom Categories (Beta)

Best Practices

When using custom categories for classification purposes, we recommend using a custom classification model and leveraging custom categories to gather initial labeled data.

When to use Categorization instead of Classification?

Categorization models work well in use cases where documents can be mapped to taxonomy nodes based on general knowledge topics discussed in the document, where these general topics have a good representation in Wikipedia. Categorization is not expected to work well in use cases that are very domain specific (with little or no representation in Wikipedia) or that require interpretation beyond general topics. For example tasks such as sentiment or emotion classification are not a good match for training Categorization models.

Use Categorization when you do not have the ability to obtain labeled data. You may also use Categorization to bootstrap labeled data for a classifier: first create a Categorization model, then apply it on a set of documents and ask human annotators to manually vet the output categories. This process results in a labeled dataset that you can then use to train a Classifier.