

► Copyright 2019 The TensorFlow Hub Authors.

Licensed under the Apache License, Version 2.0 (the "License");

↳ 1 cell hidden

► Multilingual Universal Sentence Encoder



[Run in Google Colab](#)

[View source on GitHub](#)

This notebook illustrates how to access the Multilingual Universal Sentence Encoder module and use it for sentence similarity across multiple languages. This module is an extension of the [original Universal Encoder module](#).

▼ Getting Started

This section sets up the environment for access to the Multilingual Universal Sentence Encoder Module and also prepares a set of English sentences and their translations. In the following sections, the multilingual module will be used to compute similarity *across languages*.

```
1 # Install the latest TensorFlow version compatible with tf-sentencepiece.
2 !pip3 install --quiet tensorflow==1.12.0
3 # Install TF-Hub.
4 !pip3 install --quiet tensorflow-hub
5 !pip3 install --quiet seaborn
6 # Install Sentencepiece.
7 !pip3 install --quiet tf-sentencepiece
```



More detailed information about installing Tensorflow can be found at

<https://www.tensorflow.org/install/>.

```
1 import tensorflow as tf
2 import tensorflow_hub as hub
3 import numpy as np
4 import seaborn as sns
5 import tf_sentencepiece
```

```
1 # Some texts of different lengths in different languages.
2 english_sentences = ["dog", "Puppies are nice.", "I enjoy taking long walks al
3 spanish_sentences = ["perro", "Los cachorros son agradables.", "Disfruto de d
4 german_sentences = ["Hund", "Welpen sind nett.", "Ich genieße lange Spaziergär
5 french_sentences = ["chien", "Les chiots sont gentils.", "J'aime faire de long
6 italian_sentences = ["cane", "I cuccioli sono carini.", "Mi piace fare lunghe
7 chinese_sentences = ["狗", "小狗很好.", "我喜欢和我的狗一起沿着海滩散步。"]
8 korean_sentences = ["개", "강아지가 좋다.", "나는 나의 산책을 해변을 따라 길게 산책하는 것을
9 japanese_sentences = ["犬", "子犬はいいです", "私は犬と一緒にビーチを散歩するのが好きで
```

▼ Computing Text Embeddings

We first precompute the embeddings for all of our sentences.

```
1 # The 8-language multilingual module. There are also en-es, en-de, and en-fr l
2 module_url = "https://tfhub.dev/google/universal-sentence-encoder-xling-man/1
3
4 # Set up graph.
5 g = tf.Graph()
6 with g.as_default():
7     text_input = tf.placeholder(dtype=tf.string, shape=[None])
8     xling_8_embed = hub.Module(module_url)
9     embedded_text = xling_8_embed(text_input)
10    init_op = tf.group([tf.global_variables_initializer(), tf.tables_initializer
11 g.finalize()
12
13 # Initialize session.
14 session = tf.Session(graph=g)
15 session.run(init_op)
16
17 # Compute embeddings.
18 en_result = session.run(embedded_text, feed_dict={text_input: english_sentence
19 es_result = session.run(embedded_text, feed_dict={text_input: spanish_sentence
20 de_result = session.run(embedded_text, feed_dict={text_input: german_sentence
21 fr_result = session.run(embedded_text, feed_dict={text_input: french_sentence
22 it_result = session.run(embedded_text, feed_dict={text_input: italian_sentence
23 zh_result = session.run(embedded_text, feed_dict={text_input: chinese_sentence
24 ko_result = session.run(embedded_text, feed_dict={text_input: korean_sentence
25 ja_result = session.run(embedded_text, feed_dict={text_input: japanese_sentence
```



```
WARNING:tensorflow:From /usr/local/lib/python3.6/dist-packages/tensorflow/p
Instructions for updating:
Colocations handled automatically by placer.
W0604 11:46:06.739412 140387101284224 deprecation.py:323] From /usr/local/l
Instructions for updating:
Colocations handled automatically by placer.
INFO:tensorflow:Saver not created because there are no variables in the gra
I0604 11:46:17.598348 140387101284224 saver.py:1483] Saver not created beca
```

▼ Visualize Embedding Similarity

With the sentence embeddings now in hand, we can visualize semantic similarity across different languages.

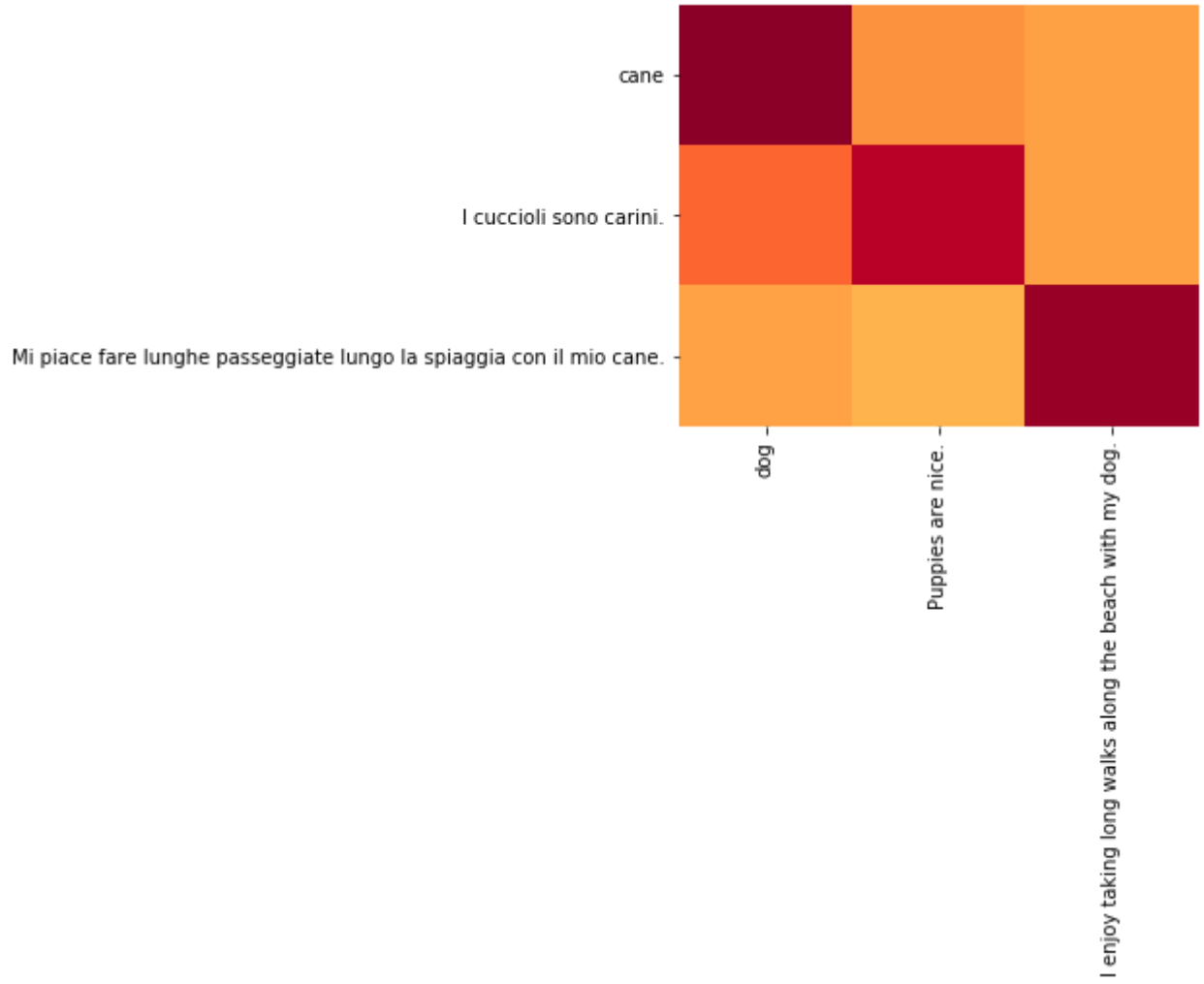
```
1 def visualize_similarity(embeddings_1, embeddings_2, labels_1, labels_2, plot_
2     corr = np.inner(embeddings_1, embeddings_2)
3     g = sns.heatmap(corr,
4                     xticklabels=labels_1,
5                     yticklabels=labels_2,
6                     vmin=0,
7                     vmax=1,
8                     cmap="YlOrRd")
9     g.set_yticklabels(g.get_yticklabels(), rotation=0)
10    g.set_title(plot_title)
```

▼ English-Italian Similarity

```
1 visualize_similarity(en_result, it_result, english_sentences, italian_sentence
```



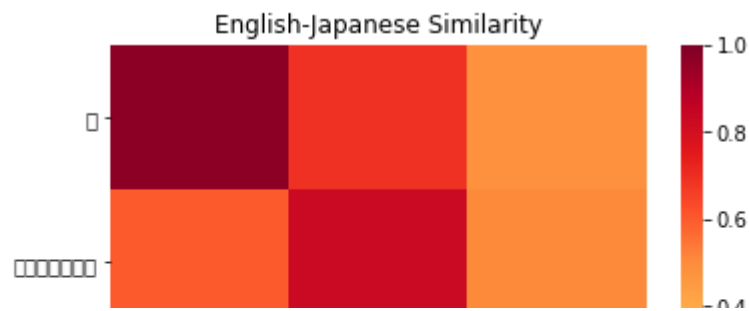
English-Italian Similarity



▼ English-Japanese Similarity

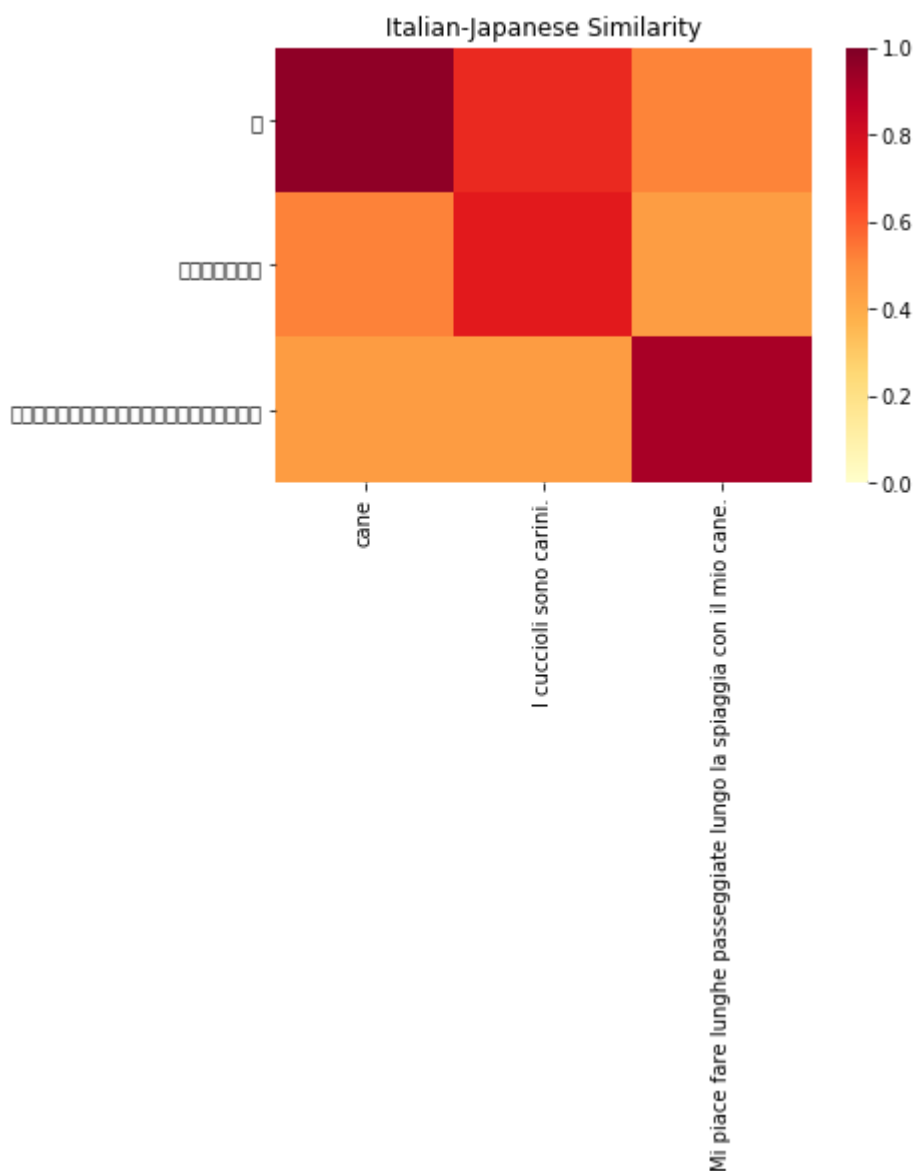
```
1 visualize_similarity(en_result, ja_result, english_sentences, japanese_sentences)
```





▼ Italian-Japanese Similarity

```
1 visualize_similarity(it_result, ja_result, italian_sentences, japanese_sentences)
```



And more...

The above examples can be extended to any language pair from **English, Spanish, German, French, Italian, Chinese, Korean, and Japanese**. Happy coding!