

```
#@title Licensed under the Apache License, Version 2.0 (the "License");
# you may not use this file except in compliance with the License.
# You may obtain a copy of the License at
#
# https://www.apache.org/licenses/LICENSE-2.0
#
# Unless required by applicable law or agreed to in writing, software
# distributed under the License is distributed on an "AS IS" BASIS,
# WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied.
# See the License for the specific language governing permissions and
# limitations under the License.
```



▼ Copyright 2019 The TensorFlow Authors.

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

```
!wget --no-check-certificate \
https://storage.googleapis.com/laurencemoroney-blog.appspot.com/sarcasm.json \
-O /tmp/sarcasm.json
```

```
import json
```

```
with open("/tmp/sarcasm.json", 'r') as f:
    datastore = json.load(f)
```

```
sentences = []
labels = []
urls = []
for item in datastore:
    sentences.append(item['headline'])
    labels.append(item['is_sarcastic'])
    urls.append(item['article_link'])
```

```
from tensorflow.keras.preprocessing.text import Tokenizer
from tensorflow.keras.preprocessing.sequence import pad_sequences
tokenizer = Tokenizer(oov_token="<OOV>")
tokenizer.fit_on_texts(sentences)
```

```
word_index = tokenizer.word_index
print(len(word_index))
print(word_index)
```



```

sequences = tokenizer.texts_to_sequences(sentences)
padded = pad_sequences(sequences, padding='post')
print(padded[0])
print(padded.shape)

```



```

--2020-09-19 14:03:38-- https://storage.googleapis.com/laurencemoroney-blog.appspot.com
Resolving storage.googleapis.com (storage.googleapis.com)... 173.194.217.128, 172.217.206.128
Connecting to storage.googleapis.com (storage.googleapis.com)|173.194.217.128|:443... cc
HTTP request sent, awaiting response... 200 OK
Length: 5643545 (5.4M) [application/json]
Saving to: '/tmp/sarcasm.json'

```

```

/tmp/sarcasm.json 100%[=====>] 5.38M --.-KB/s in 0.05s

```

```

2020-09-19 14:03:38 (108 MB/s) - '/tmp/sarcasm.json' saved [5643545/5643545]

```

```

29657

```

```

{'<OOV>': 1, 'to': 2, 'of': 3, 'the': 4, 'in': 5, 'for': 6, 'a': 7, 'on': 8, 'and': 9,
 [ 308 15115 679 3337 2298 48 382 2576 15116 6 2577 8434
    0 0 0 0 0 0 0 0 0 0 0 0
    0 0 0 0 0 0 0 0 0 0 0 0
    0 0 0 0]

```

```

(26709, 40)

```

```

--2020-09-19 14:03:38-- https://storage.googleapis.com/laurencemoroney-blog.appspot.com
Resolving storage.googleapis.com (storage.googleapis.com)... 173.194.217.128, 172.217.206.128
Connecting to storage.googleapis.com (storage.googleapis.com)|173.194.217.128|:443... cc
HTTP request sent, awaiting response... 200 OK
Length: 5643545 (5.4M) [application/json]
Saving to: '/tmp/sarcasm.json'

```

```

/tmp/sarcasm.json 100%[=====>] 5.38M --.-KB/s in 0.05s

```

```

2020-09-19 14:03:38 (108 MB/s) - '/tmp/sarcasm.json' saved [5643545/5643545]

```

```

29657

```

```

{'<OOV>': 1, 'to': 2, 'of': 3, 'the': 4, 'in': 5, 'for': 6, 'a': 7, 'on': 8, 'and': 9,
 [ 308 15115 679 3337 2298 48 382 2576 15116 6 2577 8434
    0 0 0 0 0 0 0 0 0 0 0 0
    0 0 0 0 0 0 0 0 0 0 0 0
    0 0 0 0]

```

```

(26709, 40)

```