tf.contrib.lookup.index_table_from_file

```
index_table_from_file(
    vocabulary_file=None,
    num_oov_buckets=0,
    vocab_size=None,
    default_value=-1,
    hasher_spec=tf.contrib.lookup.FastHashSpec,
    key_dtype=tf.string,
    name=None
)
```

Defined in `tensorflow/python/ops/lookup_ops.py`.

Returns a lookup table that converts a string tensor into int64 IDs.

This operation constructs a lookup table to convert tensor of strings into int64 IDs. The mapping can be initialized from a vocabulary file specified in `vocabulary_file`, where the whole line is the key and the zero-based line number is the ID.

Any lookup of an out-of-vocabulary token will return a bucket ID based on its hash if `num_oov_buckets` is greater than zero. Otherwise it is assigned the `default_value`. The bucket ID range is `[vocabulary size, vocabulary size + num_oov_buckets - 1]`.

The underlying table must be initialized by calling `tf.tables_initializer.run()` or `table.init.run()` once.

Sample Usages:

If we have a vocabulary file "test.txt" with the following content:

```
emerson
lake
palmer
```

```
features = tf.constant(["emerson", "lake", "and", "palmer"])
table = tf.contrib.lookup.index_table_from_file(
    vocabulary_file="test.txt", num_oov_buckets=1)
ids = table.lookup(features)
...
tf.tables_initializer().run()

ids.eval()  ==> [0, 1, 3, 2]  # where 3 is the out-of-vocabulary bucket
```

Args:

- `vocabulary_file`: The vocabulary filename, may be a constant scalar `Tensor`.
- `num_oov_buckets`: The number of out-of-vocabulary buckets.
- `vocab_size`: Number of the elements in the vocabulary, if known.
- `default_value`: The value to use for out-of-vocabulary feature values. Defaults to -1.
- `hasher_spec`: A `HasherSpec` to specify the hash function to use for assignation of out-of-vocabulary buckets.
- `key_dtype`: The `key` data type.

- `name` : A name for this op (optional).

Returns:

The lookup table to map a `key_dtype` `Tensor` to index `int64` `Tensor` .

Raises:

- `ValueError` : If `vocabulary_file` is not set.
- `ValueError` : If `num_oov_buckets` is negative or `vocab_size` is not greater than zero.

Stay Connected

Blog

GitHub

Twitter

Support

Issue Tracker

Release Notes

Stack Overflow

English

Terms | Privacy