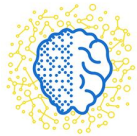
Named Entity Recognition with DeepPavlov library

Mikhail Arkhipov
Laboratory of Neural Systems and Deep Learning
MIPT

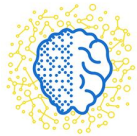
Named Entity Recognition



Named Entities:

- names
- organizations
- locations
- time expressions
- quantities
- monetary values

Example

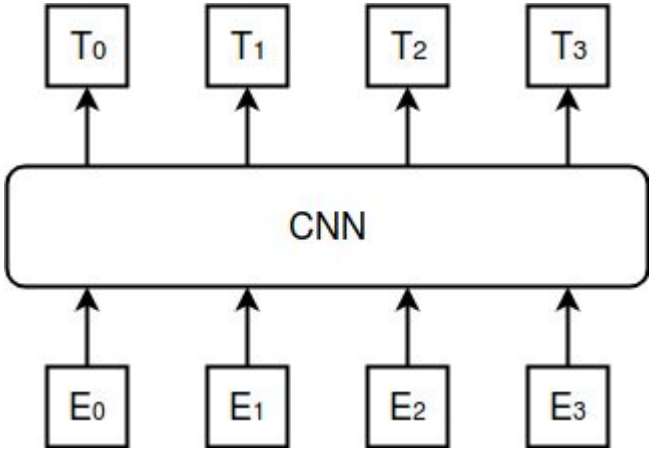


Zdzisław Beksiński was born in Sanok Poland

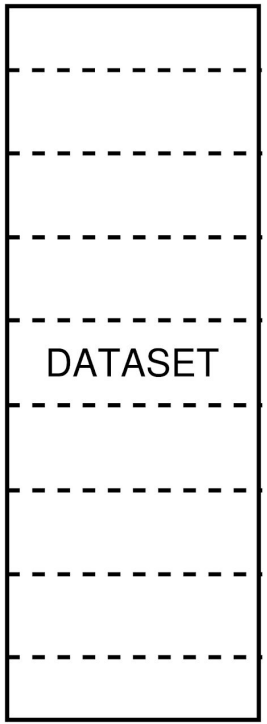
B-PER I-PER O O O B-LOC B-LOC

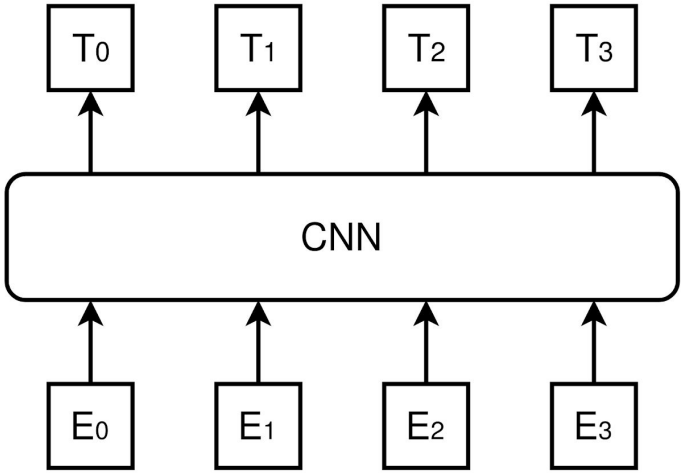
- Person
- Location

Architecture



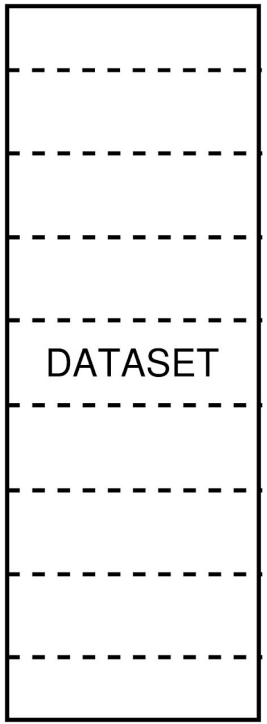
Dataset

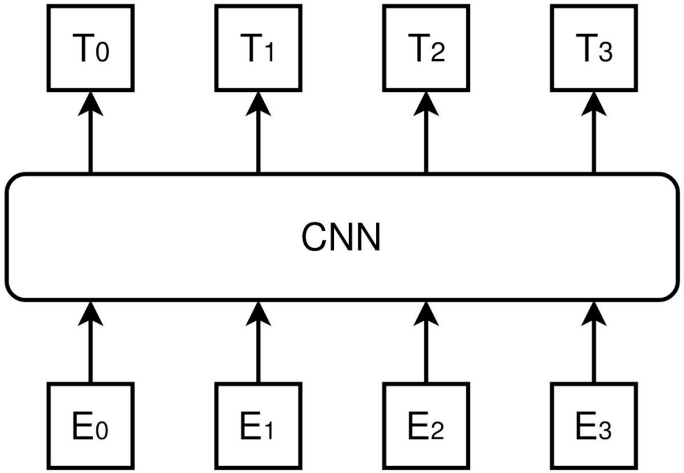




Dataset

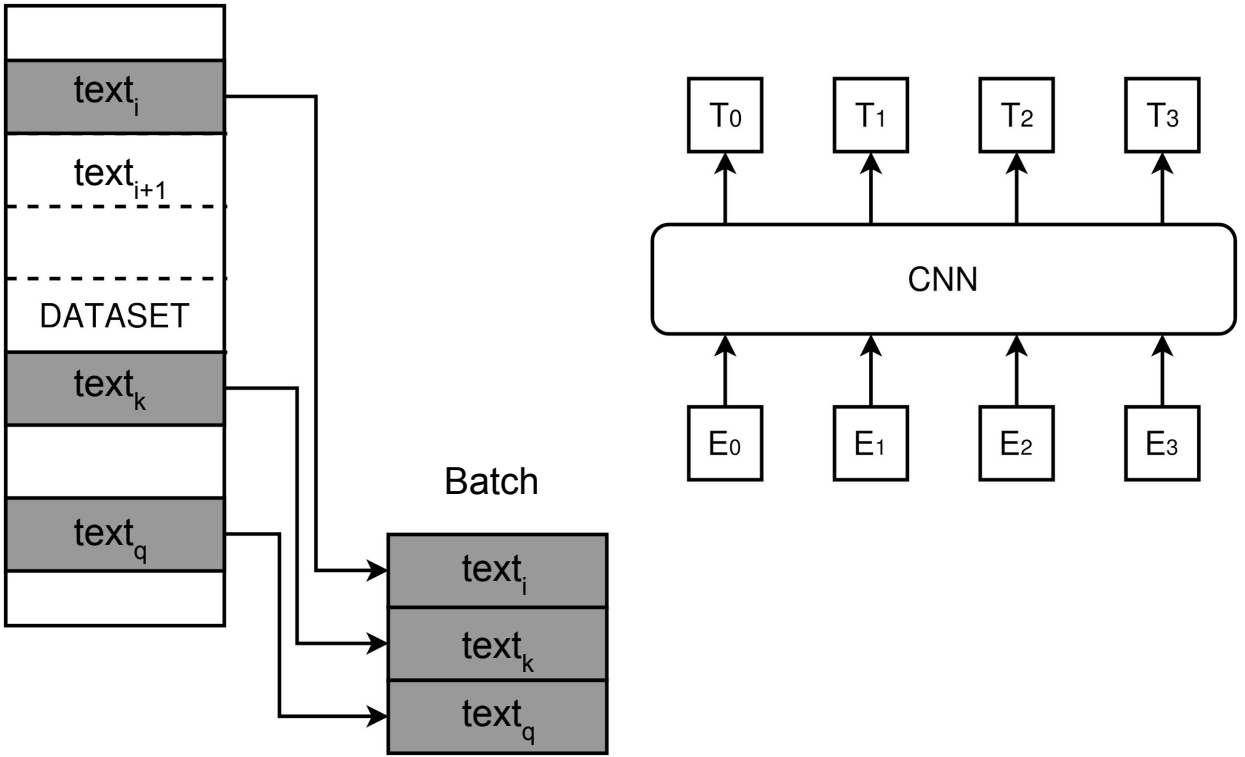
```
dp.dataset_readers.conll2003_reader
```





Batch Generator

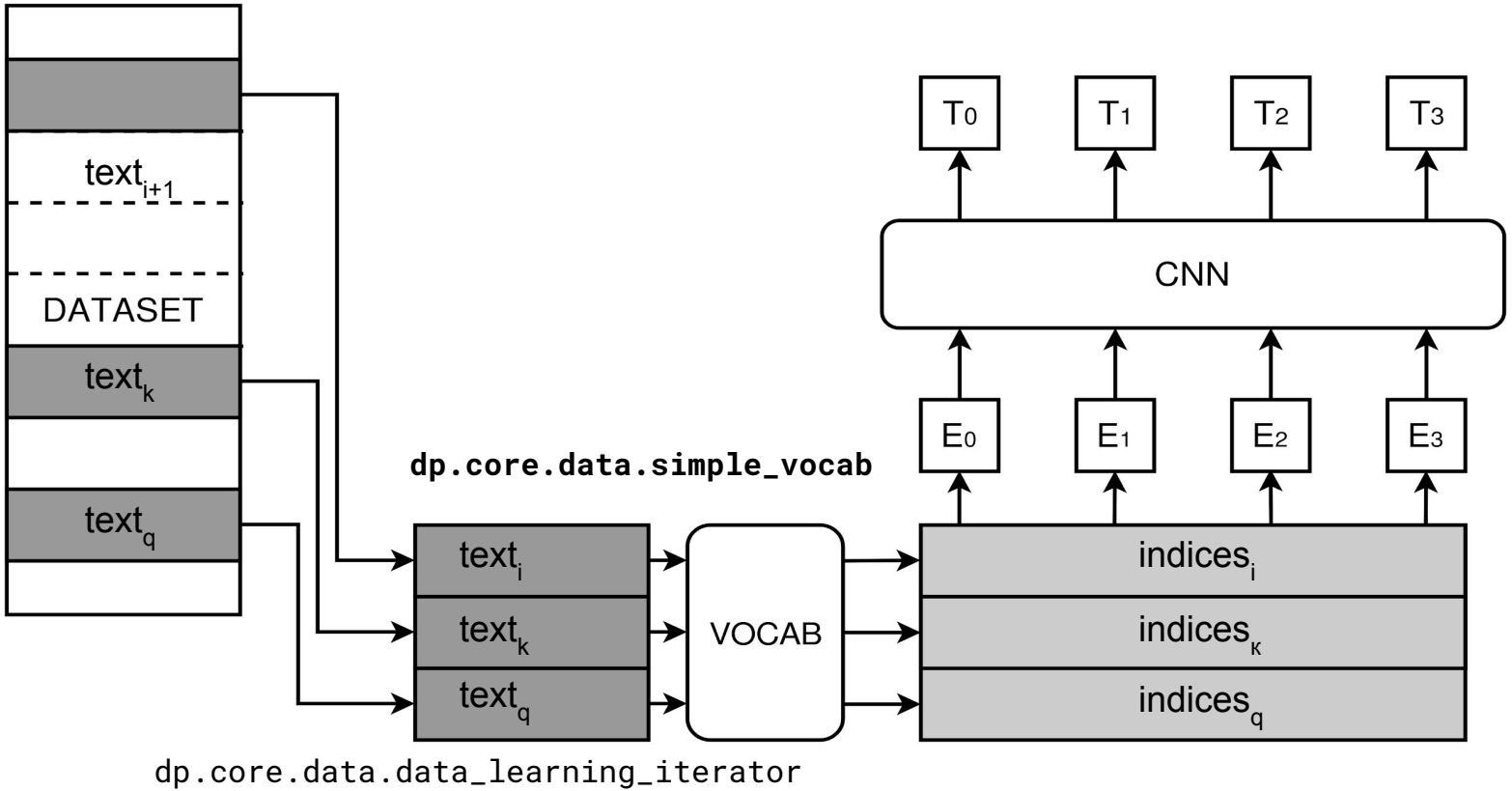
`dp.dataset_readers.conll2003_reader`



`dp.core.data.data_learning_iterator`

Vocabulary

`dp.dataset_readers.conll2003_reader`

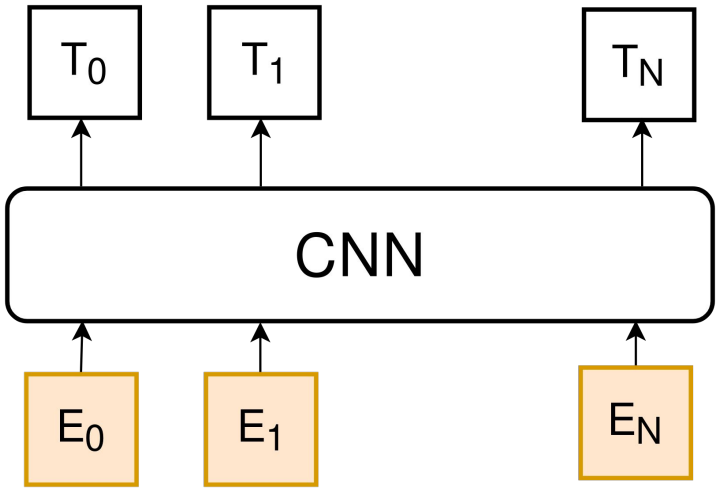


Embeddings

```python
def get_embeddings(indices, vocabulary_size, emb_dim):
    # Initialize the random gaussian matrix with dimensions [vocabulary_size, embedding_dimension]
    # The **VARIANCE** of the random samples must be 1 / embedding_dimension

    # YOUR CODE HERE

    emb_mat = tf.Variable(emb_mat, trainable=True, dtype=tf.float32)
    emb = tf.nn.embedding_lookup(emb_mat, indices)
    return emb
```

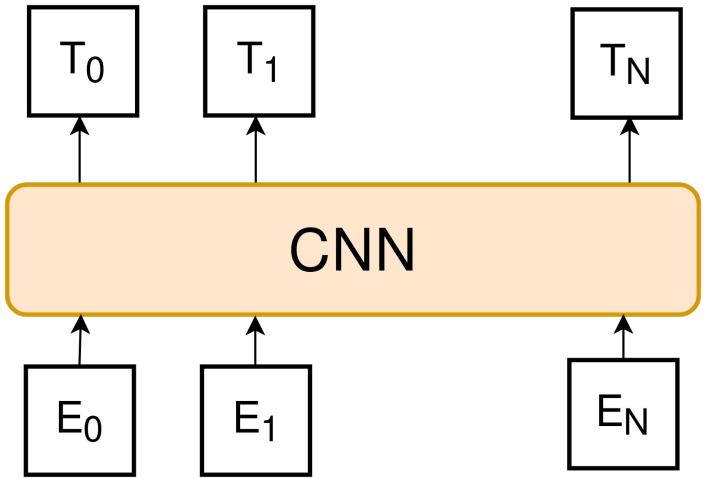


Convolutional Neural Network

```python
def conv_net(units, n_hidden_list, cnn_filter_width, activation=tf.nn.relu):
    # Use activation(units) to apply activation to units

    ########################################
    ########## YOUR CODE HERE ############
    ########################################

    return units
```



Loss function

```
def masked_cross_entropy(logits, label_indices, number_of_tags, mask):

    ########################################
    ########## YOUR CODE HERE ##############
    ########################################

    return loss
```