

### Training with bert-base-uncased:

No explicit dev set, all sequences were shuffled.

Epochs: **3**

Duration: **ca. 16 min**

Sequence length: **64 (words), 128 (tokens)**

Batch size: **16**

Train samples: **21696**

Test/dev samples: **2411**

	Precision	Recall	F1-score	support
<i>BOS</i>	<b>0.79728</b>	<b>0.79925</b>	<b>0.79826</b>	<b>7706</b>
<i>O</i>	<b>0.98945</b>	<b>0.98932</b>	<b>0.98932</b>	<b>146598</b>
<i>accuracy</i>			<b>0.97983</b>	<b>154304</b>
<i>Macro avg</i>	<b>0.89336</b>	<b>0.89428</b>	<b>0.89382</b>	<b>154304</b>
<i>Weighted avg</i>	<b>0.97985</b>	<b>0.97983</b>	<b>0.97984</b>	<b>154304</b>

Confusion matrix:

True	BOS	<b>6159</b>	<b>1547</b>
	O	<b>1566</b>	<b>145032</b>
		BOS	O
		Predicted	