

One-hot Encoding

Tokenization → { Sentence
word
N-gram

"Should I learn NLP"

Step 01. Lower-case

Step 02. Ascending Sorting (A-Z)

Step 03. Give a label

Step 04. Transforming to a binary vector

S1) "should i learn nlp"

S2) "i learn nlp should"

S3) " 1 2 3 0

S4) $\begin{bmatrix} [1, ., ., .] & \# \text{ should} \\ [. , 1, ., .] & \# i \\ [. , ., 1, .] & \# \text{ learn} \\ [., ., ., 1] & \# \text{ nlp} \end{bmatrix}$