

- 4) Write a python code for implementing the “Byte-pair Encoding” tokenization algorithm.

- 5) Mention a pair of words having:
 - a. Same lemmas and same stems
 - b. Same lemmas and different stems
 - c. Different lemmas and same stems
 - d. Different lemmas and different stems