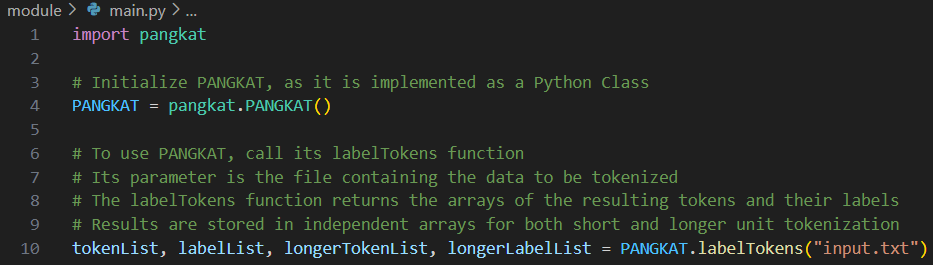
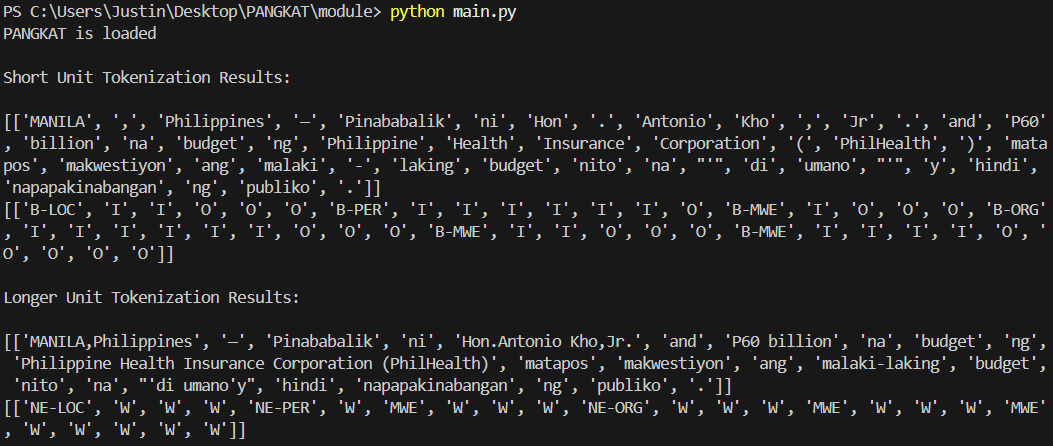
**Manual for PANGKAT: A Dedicated Tokenizer for the Tagalog Language**

PANGKAT was designed to be implemented as a Python module, which can be imported into more advanced Tagalog NLP applications as its tokenizer.



For instance, the results of PANGKAT are as follows for tokenizing the following sentence:  
  
“MANILA, Philippines — Pinababalik ni Hon. Antonio Kho, Jr. and P60 billion na budget ng Philippine Health Insurance Corporation (PhilHealth) matapos makwestiyon ang malaki-laking budget nito na 'di umano'y hindi napapakinabangan ng publiko.”



For better visualization, the following images present the results using PANGKAT’s graphical user interface for both short and longer unit tokenization, respectively:

