This python code is functioned as a vocabulary analysis machine of JSON file exported from Zeppelin, It has four main function,

* Generate different paragraphs and make them executable separately
* Generate a bash file to realize the sequence all the script run on Zeppelin
* Add the procedure of detecting the definition of functions and variables to solve the problem of the definition and call of function or variable take place at different paragraphs.
* Detect and illustrate the dependency between python script by graph,

According to the definition of JSON file, the part we need to use are included in the section of “text” and “id”, the former one is the content of each paragraph and the later one is the label of each paragraph according to the time label it created.

The main function of the script is aimed to the whole JSON file, when read one whole part of paragraph, it will pass this part to the process function, to realize the function mentioned before.

The process of most language paragraph is almost the same. For most characters, they would be passed to the print list after read in. But there are extra process for some special characters, such as (“), which is the end symbol of a text, as well as some control characters such as “\n” and “\t”.

This procedure only guarantees the accord of the script exported the same as that in Zeppelin. There may be some other steps to make the script executable. For example, the shell script can be generated and ran without and process. However for python code, main function and some control sentences need to be added to the script.