## **Using Sylbreak with Jupyter Notebook**

ဒီ Jupyter Notebook က GitHub မှာ ကျွန်တော်တင်ပေးထားတဲ့ Sylbreak Python ပရိုဂရမ် <a href="https://github.com/ye-kyaw-thu/sylbreak/blob/master/python/sylbreak.py">https://github.com/ye-kyaw-thu/sylbreak/blob/master/python/sylbreak/blob/master/python/sylbreak.py</a>) ကို Jupyter Notebook, Python 3 Kernel မှာ copy ကူးပြီးသုံးတဲ့အခါမှာ တွေ့နိုင်တဲ့ error နဲ့ solution ကို နမူနာအနေနဲ့ ပြသထားတာ ဖြစ်ပါတယ်။

အောက်ပါ Python code က sylbreak.py ထဲက ယူထားတာဖြစ်ပါတယ်။ ဒီနေရာမှာတော့ မြန်မာစာကို syllable unit အနေနဲ့ ဖြတ်ပေးဖို့အတွက် လိုအပ်တဲ့ အပိုင်းကိုပဲ ယူထားတာဖြစ်ပြီးတော့၊ function တစ်ခုအနေနဲ့ ဆောက်ထားတာပါ။

statement အသစ်ရေးထားတဲ့ line = re.sub(ur"\s+","", line) ဆိုတာကတော့ function ကို pass လုပ်လိုက်တဲ့ စာကြောင်းကို space ဖယ်တဲ့ လိုင်းပါ။

```
In [1]:
             # Syllable Segmentation for Myanmar language
             import re
          3
            myConsonant = ur"თ-თ"
             enChar = ur"a-zA-Z0-9"
          6 otherChar = ur"ဣဤဉဦီသြေသြော်ဿ၌၍၏ - ၉။၊ ! - / : -@[ - ` { -~\s"
             ssSymbol = ur'o
             ngaThat = ur'δ<sup>1</sup>
          9
             aThat = ur'&'
         10
         11
            #Regular expression pattern for Myanmar syllable breaking
         12
            #*** a consonant not after a subscript symbol AND a consonant is not
         13
             BreakPattern = re.compile(ur"((?<!" + ssSymbol + ur")["+ myConsonant
         14
         15
             def sylbreak(line):
                    line = re.sub(ur"\s+","", line)
         16
         17
                    line = BreakPattern.sub(ur" " + ur"\1", line)
         18
                    return line
```

File "<ipython-input-1-alf5ba8fedd0>", line 4 myConsonant = ur"თ-თ"

SyntaxError: invalid syntax

အထက်ပါ cell ကို "Ctrl+Enter" (သို့) "Shift+Enter" နဲ့ run တဲ့အခါမှာ error ပေးပါလိမ့်မယ်။ အကြောင်းအရင်းက string တွေရဲ့ ရှေ့မှာ ဖြည့်ရေးထားတဲ့ "ur" ကြောင့်ပါ။ Python 3.4 မှာ u (Unicode) နဲ့ r (raw text) ကို တစ်လုံးချင်းစီသုံးတာကို support လုပ်ပေမဲ့၊ "ur" ဆိုပြီးတွဲသုံးတာကို support မလုပ်တော့လို့ပါ။

အဲဒါကြောင့် "ur" အစား "r" တစ်ခုတည်းပဲ သုံးပါ။ အဲဒါဆိုရင် အလုပ်လုပ်ပေးပါလိမ့်မယ်။ အောက်ပါ code က run လို့ရအောင်ပြင်ရေးထားတဲ့ Python code ပါ။

```
# Syllable Segmentation for Myanmar language
In [2]:
           2
              import re
           3
           4
             myConsonant = "თ-თ"
           5
              enChar = "a-zA-Z0-9"
             otherChar = "ဣဤဥဦီဩေသြော်ဿ၌၍၏၀-၉။ ! - / : -@[ - ` { -~\s" ssSymbol = '့ ်
           7
              ngaThat = '&
           8
           9
              aThat = '\delta'
          10
          11
              #Regular expression pattern for Myanmar syllable breaking
          12
              #*** a consonant not after a subscript symbol AND a consonant is not
         13
              BreakPattern = re.compile(r"((?<!" + ssSymbol + r")["+ myConsonant +
         14
         15
              def sylbreak(line):
         16
                      line = re.sub(r"\s+","", line)
                      line = BreakPattern.sub(r" + r"\1", line)
          17
         18
                      return line
          1 # sylbreak function ကိုခေါ်သုံးကြည့်ရအောင်
In [3]:
           3 sylbreak("မြန်မာစာသည် တို့စာ။ တို့စာကို သုတေသန လုပ်ပါ။")
\mathsf{Out}[3]: 'မြန် မာ စာ သည် တို့ စာ ။ တို့ စာ ကို သု တေ သ န လုပ် ပါ ။'
```

## Reference

https://stackoverflow.com/questions/26063899/python-version-3-4-does-not-support-a-ur-prefix (https://stackoverflow.com/questions/26063899/python-version-3-4-does-not-support-a-ur-prefix)

sylbreak နဲ့ ပတ်သတ်ပြီးအသေးစိတ်သိချင်သူများအတွက် လင့်(ခ်) (https://github.com/ye-kyaw-thu/sylbreak/)