

Top 10 Longest word in Venmurasu Novel

- 1.வெடிபடுமண்டலத்திருளலைவெடிபடநடமிடுதுடியெழுகடியொலிதாளம் (32)
- 2.சர்வகல்விதமேவாஹம்சர்வகல்விதமேவாஹம் (22)
- 3.இருந்துகொண்டிருந்தவர்களுக்குமட்டும்தான் (22)
- 4.நான்யாஸ்திசனாதனம்சர்வகல்விதமேவாஹம் (21)
- 5.அலங்கரிக்கப்பட்டிருக்கவேண்டுமல்லவா (21)
- 6.உக்ரசேனன்சத்ருஞ்சயன்வியாஹ்ரதந்தன் (21)
- 7.நினைவுபடுத்திக்கொள்ளவேண்டியிருக்கிறதே (20)
- 8.பனங்கொட்டைத்துருவலைப்போட்டிருந்தனர் (20)
- 9.இறந்தவர்களுக்கும்தீராச்சினத்துடன் (20)
- 10.ஆண்களில்லாமலாகிவிட்டிருக்கின்றனர் (19)

First Page:

<https://venmurasu.in/2014/01/01/%e0%ae%b5%e0%af%86%e0%ae%a3%e0%af%8d%e0%ae%ae%e0%af%81%e0%ae%b0%e0%ae%9a%e0%af%81-%e0%ae%a8%e0%af%82%e0%ae%b2%e0%af%8d-%e0%ae%92%e0%ae%a9%e0%af%8d%e0%ae%b1%e0%af%81/>

Last Page:

<https://venmurasu.in/2020/07/16/%e0%ae%a8%e0%af%82%e0%ae%b2%e0%af%8d-%e0%ae%87%e0%ae%b0%e0%af%81%e0%ae%aa%e0%ae%a4%e0%af%8d%e0%ae%a4%e0%ae%bf%e0%ae%af%e0%ae%be%e0%ae%b1%e0%af%81-%e0%ae%ae%e0%af%81%e0%ae%a4%e0%ae%b2%e0%ae%be-16/>

Perquisites:

- pip install beautifulsoup4
- pip install Open-Tamil
- pip install requests

Code:

```
import requests
from bs4 import BeautifulSoup
from tamil import utf8
from itertools import count
import winsound

URL =
'https://venmurasu.in/2014/01/01/%e0%ae%b5%e0%af%86%e0%ae%a3%e0%af%8d%e0%ae%ae%e0%
af%81%e0%ae%b0%e0%ae%9a%e0%af%81-%e0%ae%a8%e0%af%82%e0%ae%b2%e0%af%8d-
%e0%ae%92%e0%ae%a9%e0%af%8d%e0%ae%b1%e0%af%81/'
final=[]
new=[]
while True:
    f = open("output.txt", "w",encoding="utf-8")
    page = requests.get(URL)

    soup = BeautifulSoup(page.content, 'html.parser')
    job_elems = soup.find('div', class_='entry-content')
    try:
        href=soup.find('a',rel="next").attrs
    except:
        print('completed')
        break
    #print(href)

    result=utf8.get_words(job_elems.text)
    for fb in final:
        result.append(fb)
    for a in result:
        a=a.replace("'",")")
        a=a.replace('"',"")
        a=a.replace('!',",10000000000")
        a=a.replace('.',",10000000000")
        a=a.replace(',',",10000000000")
        a=a.replace('?',",10000000000")
        a=a.replace('ஏற்றப்படுகின்றது',"10000000000")
        if(utf8.all_tamil(a) ):
            if a not in new:
                new.append(a)
                #f.write(a)
                #f.write("\n")

    else:
        a=""

    cnt=count()
```

```
final=sorted(new, key = lambda w : (len(utf8.get_letters(w)), next(cnt)),reverse = True)[:10]
print(final)
URL=href['href']
for fa in final:
    f.write(fa)
    f.write('\t\t')
    f.write(str(len(utf8.get_letters(fa))))
    f.write('\n')
    #f.write(href['href'])
    f.write('\n')
f.close()
winsound.Beep(2000, 1000)
```