python-wordcloud-example

April 20, 2017

```
In [1]: from konlpy.tag import Kkma
        from konlpy.utils import pprint
        from IPython.display import Image
        %matplotlib inline

In [2]: #import jpype

        #jpype.startJVM(jpype.getDefaultJVMPath())

In [3]: kkma = Kkma()

In [15]: pprint(kkma.sentences(u' .  ?   .'))

[  .,
   ?,
   .]


In [16]: pprint(kkma.nouns(u'          .'))

[,
 ,
 ,
 ,
 ,
 ,
 ,
 ,
 ]


In [20]: pprint(kkma.pos(u'  .     !^^'))

[(, NNG),
 (, NNG),
 (, JX),
 (, NNG),
 (, XSA),
```

```
 (, ETD),
 (, NNG),
 (., SF),
 (, NP),
 (, JX),
 (, MAG),
 (, NNG),
 (, VCP),
 (, ETD),
 (, NNG),
 (, JKO),
 (, VV),
 (, ECD),
 (, VV),
 (, EFN),
 (!, SF),
 (^^, EMO)]
```

```
In [33]: pprint(kkma.pos(u'4   .'))
```

```
[(4, NR),
 (, NNM),
 (, NNG),
 (, NNG),
 (, JKS),
 (, VV),
 (, EPT),
 (, EFN),
 (., SF)]
```

```
In [28]: #! /usr/bin/python2.7
         # -*- coding: utf-8 -*-

         from collections import Counter
         import urllib
         import random
         import webbrowser

         from konlpy.tag import Hannanum
         from lxml import html
         import pytagcloud # requires Korean font support
         import sys

         if sys.version_info[0] >= 3:
             urlopen = urllib.request.urlopen
         else:
```

```
        urlopen = urllib.urlopen


    r = lambda: random.randint(0,255)
    color = lambda: (r(), r(), r())

    def get_bill_text(billnum):
        url = 'http://pokr.kr/bill/%s/text' % billnum
        response = urlopen(url).read().decode('utf-8')
        page = html.fromstring(response)
        text = page.xpath(".//div[@id='bill-sections']/pre/text()")[0]
        return text

    def get_tags(text, ntags=50, multiplier=10):
        h = Hannanum()
        nouns = h.nouns(text)
        count = Counter(nouns)
        return [{ 'color': color(), 'tag': n, 'size': c*multiplier }\
                    for n, c in count.most_common(ntags)]

    def draw_cloud(tags, filename, fontname='Nanum Gothic', size=(800, 600)):
        pytagcloud.create_tag_image(tags, filename, fontname=fontname, size=size)
        webbrowser.open(filename)
```

In [29]:
```
bill_num = '1904882'
text = get_bill_text(bill_num)
tags = get_tags(text)
```

In [30]:
```
#print(tags)
```

```
[{'color': (200, 98, 45), 'tag': u'\ub3d9\ubb3c', 'size': 80}, {'color': (71, 178, 30), 'tag':
```

In [31]:
```
draw_cloud(tags, 'wordcloud.png')
```

In [32]:
```
Image('wordcloud.png')
```

Out[32]:


판매·전시·전달·상영
홍보
굶주림
표시
제8조
의식
것
약
고통
신설
이
게재
제외
인터넷
다음
의
사
~
촬영
등
호
12
학대
①
제8조제5항
④
현행법
일부개정법률안
위반

In [ ]: