

```
1 !sudo apt-get install -y fonts-nanum
2 !sudo fc-cache -fv
3 !rm ~/.cache/matplotlib -rf
```

```
Reading package lists... Done
Building dependency tree
Reading state information... Done
The following package was automatically installed and is no longer required:
  libnvidia-common-460
Use 'sudo apt autoremove' to remove it.
The following NEW packages will be installed:
  fonts-nanum
0 upgraded, 1 newly installed, 0 to remove and 20 not upgraded.
Need to get 9,604 kB of archives.
After this operation, 29.5 MB of additional disk space will be used.
Get:1 http://archive.ubuntu.com/ubuntu bionic/universe amd64 fonts-nanum all 20170925-1 [9,604 k
Fetched 9,604 kB in 1s (8,265 kB/s)
debconf: unable to initialize frontend: Dialog
debconf: (No usable dialog-like program is installed, so the dialog based frontend cannot be use
debconf: falling back to frontend: Readline
debconf: unable to initialize frontend: Readline
debconf: (This frontend requires a controlling tty.)
debconf: falling back to frontend: Teletype
dpkg-preconfigure: unable to re-open stdin:
Selecting previously unselected package fonts-nanum.
(Reading database ... 157604 files and directories currently installed.)
Preparing to unpack .../fonts-nanum_20170925-1_all.deb ...
Unpacking fonts-nanum (20170925-1) ...
Setting up fonts-nanum (20170925-1) ...
Processing triggers for fontconfig (2.12.6-0ubuntu2) ...
/usr/share/fonts: caching, new cache contents: 0 fonts, 1 dirs
/usr/share/fonts/truetype: caching, new cache contents: 0 fonts, 3 dirs
/usr/share/fonts/truetype/humor-sans: caching, new cache contents: 1 fonts, 0 dirs
/usr/share/fonts/truetype/liberation: caching, new cache contents: 16 fonts, 0 dirs
/usr/share/fonts/truetype/nanum: caching, new cache contents: 10 fonts, 0 dirs
/usr/local/share/fonts: caching, new cache contents: 0 fonts, 0 dirs
/root/.local/share/fonts: skipping, no such directory
/root/.fonts: skipping, no such directory
/var/cache/fontconfig: cleaning cache directory
/root/.cache/fontconfig: not cleaning non-existent cache directory
/root/.fontconfig: not cleaning non-existent cache directory
fc-cache: succeeded
```

▼ 1. 문자열 처리

```
1 text = '챗봇(chatbot) 혹은 채터봇(chatterbot)은 음성이나 문자를 통한 인간과의 대화를 통해서 특정한 :

1 text_upper = text.upper()
```

```
2 text_upper
```

'챗봇(Chatbot) 혹은 채터봇(Chatterbot)은 음성이나 문자를 통한 인간과의 대화를 통해서 특정
한 작업을 수행하도록 제작된 컴퓨터 프로그래밍이다. 토크봇(talkbot), 채터박스(chatbox) 혹은

```
1 text_lower = text.lower()
```

```
2 text_lower
```

'챗봇(chatbot) 혹은 채터봇(chatterbot)은 음성이나 문자를 통한 인간과의 대화를 통해서 특정
한 작업을 수행하도록 제작된 컴퓨터 프로그래밍이다. 토크봇(talkbot), 채터박스(chatterbox) 혹은

```
1 text.find('컴퓨터')
```

```
75
```

```
1 pos = text.find('컴퓨터')
```

```
2 text[pos : pos+len('컴퓨터 프로그램')]
```

```
'컴퓨터 프로그램'
```

```
1 '인공지능' in text
```

```
False
```

```
1 len(text)
```

```
137
```

```
1 text.count('bot')
```

```
4
```

▼ 2. WordCloud 생성

```
1 from google.colab import files
```

```
2 files.upload()
```

```

\xeb\x84\xec\x84\x9c \xec\x81\x97\xeb\x84\x87\r\n\xec\x81\x97\xeb\x84\x87(chatbot)
\xed\x98\xb9\xec\x9d\x80 \xec\x81\x84\xed\x84\xb0\xeb\x84\x87(chatterbot)\xec\x9d\x80
\xec\x9d\x8c\xec\x84\xb1\xec\x9d\xb4\xeb\x82\x98 \xeb\xac\xb8\xec\x9e\x90\xeb\xa5\xbc
\xed\x86\xb5\xed\x95\x9c \xec\x9d\xb8\xea\xb0\x84\xea\xb3\xbc\xec\x9d\x98
\xeb\x8c\x80\xed\x99\x94\xeb\xa5\xbc \xed\x86\xb5\xed\x95\xb4\xec\x84\x9c
\xed\x8a\xb9\xec\xa0\x95\xed\x95\x9c \xec\x9e\x91\xec\x97\x85\xec\x9d\x84
\xec\x88\x98\xed\x96\x89\xed\x95\x98\xeb\x8f\x84\xeb\xa1\x9d
\xec\xa0\x9c\xec\x9e\x91\xeb\x90\x9c \xec\xbb\xb4\xed\x93\xa8\xed\x84\xb0
\xed\x94\x84\xeb\xa1\x9c\xea\xb7\xb8\xeb\x9e\xa8\xec\x9d\xb4\xeb\x8b\xa4.
\xed\x86\xa0\xed\x81\xac\xeb\x84\x87(talkbot),
\xec\x81\x84\xed\x84\xb0\xeb\xb0\x95\xec\x8a\xa4(chatterbox) \xed\x98\xb9\xec\x9d\x80
\xea\xb7\xb8\xeb\x83\xa5 \xeb\x84\x87(bot)\xeb\x9d\xbc\xea\xb3\xa0\xeb\x8f\x84
\xed\x95\x9c\xeb\x8b\xa4.\r\n\r\n\xec\x9d\xb8\xea\xb0\x84\xec\x9d\xb4
\xec\x82\xac\xec\x9a\xa9\xed\x95\x98\xeb\x8a\x94 \xec\x96\xb8\xec\x96\xb4\xeb\xa5\xbc
\xec\x9d\xb4\xed\x95\xb4\xed\x95\xb4\xec\x84\x9c \xeb\x8c\x80\xed\x99\x94\xeb\xa5\xbc
\xec\xa7\x84\xed\x96\x89\xed\x95\x98\xeb\x8a\x94
\xec\x9e\x90\xec\x97\xb0\xec\x96\xb8\xec\x96\xb4\xec\xb2\x98\xeb\xa6\xac
\xea\xb8\xb0\xec\x88\xa0\xec\x9d\x84 \xec\x8b\xac\xea\xb0\x81\xed\x95\x98\xea\xb2\x8c
\xec\xa0\x81\xec\x9a\xa9\xed\x95\x98\xeb\x8a\x94
\xec\x96\xb8\xec\x96\xb4\xec\x9d\xb4\xed\x95\xb4 \xeb\xb0\xa9\xec\x8b\x9d,
\xec\x9e\x85\xeb\xa0\xa5\xeb\xb0\x9b\xec\x9d\x80 \xeb\xa7\x90\xec\x97\x90\xec\x84\x9c
\xed\x8a\xb9\xec\xa0\x95 \xeb\x8b\xa8\xec\x96\xb4\xeb\x82\x98
\xec\x96\xb4\xea\xb5\xac\xeb\xa5\xbc \xea\xb2\x80\xec\xb6\x9c\xed\x95\x98\xec\x97\xac
\xea\xb7\xb8\xec\x97\x90 \xeb\xa7\x9e\xeb\x8a\x94 \xeb\xaf\xb8\xeb\xa6\xac
\xec\xa4\x80\xeb\xb9\x84\xeb\x90\x9c \xec\x9d\x91\xeb\x8b\xb5\xec\x9d\x84
\xec\xb6\x9c\xeb\xa0\xa5\xed\x95\x98\xeb\x8a\x94 \xea\xb2\x80\xec\x83\x89
\xeb\xb0\xa9\xec\x8b\x9d,\r\n\r\n\xea\xb7\xb8\xeb\xa6\xac\xea\xb3\xa0
\xea\xb0\x81\xeb\xb3\xb8\xec\x9d\x84 \xeb\xaf\xb8\xeb\xa6\xac
\xeb\xa7\x8c\xeb\x93\xa4\xea\xb3\xa0 \xea\xb0\x81\xeb\xb3\xb8\xec\x97\x90
\xeb\x94\xb0\xeb\x9d\xbc\xec\x84\x9c \xec\x82\xac\xec\x9a\xa9\xec\x9e\x90\xec\x9d\x98
\xec\x9e\x85\xeb\xa0\xa5\xec\x97\x90 \xeb\x8c\x80\xed\x95\x9c
\xeb\x8f\x99\xec\x9e\x91\xea\xb3\xbc \xea\xb0\x81\xeb\xb3\xb8\xec\x97\x90
\xec\x9e\x88\xeb\x8a\x94 \xec\x9d\x91\xeb\x8b\xb5\xec\x9d\x84
\xec\xb6\x9c\xeb\xa0\xa5\xed\x95\x98\xeb\x8a\x94 \xea\xb0\x81\xeb\xb3\xb8
\xeb\xb0\xa9\xec\x8b\x9d\xec\x9d\xb4
\xec\x9e\x88\xeb\x8b\xa4.\r\n\r\n\xeb\x84\xa4\xec\x9d\xb4\xeb\xb2\x84\xed\x81\xb4\xeb\x8c
\xec\xb9\xb4\xec\xb9\xb4\xec\x98\xa4\xed\x86\xa1, LINE, \xec\x8a\xac\xeb\x9e\x99,
\xed\x8e\x98\xec\x9d\xb4\xec\x8a\xa4\xeb\xb6\x81 \xeb\xa9\x94\xec\x8b\xa0\xec\xa0\x80,
\xec\x8a\xa4\xeb\x83\x87\xec\xb9\x98\xeb\xb4\x87 \xeb\x93\xb1\xec\x97\x90\xec\x84\x9c
\xec\xb1\x97\xeb\x84\x87\xec\x9d\x84 \xea\xb5\xac\xeb\x8f\x99\xed\x95\xa0 \xec\x88\x98
\xec\x9e\x88\xeb\x8a\x94 API\xeb\xa5\xbc
\xec\xa0\x9c\xea\xb3\xb5\xed\x95\x9c\xeb\x8b\xa4.\r\n\r\n\xea\xb0\x9c\xeb\xb0\x9c\r\n\xec
\xec\x97\xad\xec\x82\xac \xec\x84\x88\xea\xb8\xb0\xec\x9d\x98
\xec\x81\x84\xed\x84\xb0\xeb\xb4\x87\xec\x9d\x80 ELIZA(1966\xeb\x85\x84)\xec\x99\x80
PARRY(1972\xeb\x85\x84)\xec\x9d\xb4\xeb\x8b\xa4.[1][2][3][4] \xeb\x8d\x94
\xec\xb5\x9c\xea\xb7\xbc\xec\x9d\x98 \xec\xa0\x80\xeb\xaa\x85\xed\x95\x9c
\xed\x94\x84\xeb\xa1\x9c\xea\xb7\xb8\xeb\x9e\xa8\xeb\x93\xa4\xec\x97\x90\xeb\x8a\x94
A.L.I.C.E., Jabberwacky, D.U.D.E\xea\xb0\x80 \xec\x9e\x88\xeb\x8b\xa4.
2017\xeb\x85\x84, \xec\x9d\xb4\xec\x8a\xa4\xeb\x9d\xbc\xec\x97\x98
\xed\x9a\x8c\xec\x82\xac SnatchBot\xec\x9d\x80 \xea\xb0\x90\xec\xa0\x95\xec\x9d\x84
\xeb\xb6\x84\xec\x84\x9d\xed\x95\x98\xeb\x8a\x94 \xeb\xb4\x87\xec\x9d\x84

```

```
1 !pip install koNLPy
```

```

Looking in indexes: https://pypi.org/simple, https://us-python.pkg.dev/colab-wheels/public/simple
Collecting koNLPy
  Downloading koNLPy-0.6.0-py2.py3-none-any.whl (19.4 MB)
    |████████████████████████████████████████| 19.4 MB 1.4 MB/s
Collecting JPype1>=0.7.0
  Downloading JPype1-1.4.0-cp37-cp37m-manylinux_2_5_x86_64.manylinux1_x86_64.whl (453 kB)
    |████████████████████████████████████████| 453 kB 56.9 MB/s
Requirement already satisfied: numpy>=1.6 in /usr/local/lib/python3.7/dist-packages (from koNLPy)
Requirement already satisfied: lxml>=4.1.0 in /usr/local/lib/python3.7/dist-packages (from koNLPy)
Requirement already satisfied: typing-extensions in /usr/local/lib/python3.7/dist-packages (from koNLPy)
Installing collected packages: JPype1, koNLPy
Successfully installed JPype1-1.4.0 koNLPy-0.6.0

```

```

1 from wordcloud import WordCloud
2 from wordcloud import STOPWORDS
3 import matplotlib.pyplot as plt
4
5 from PIL import Image
6 import numpy as np

```

```

1 f = open('./챗봇.txt', 'r', encoding='UTF8')
2 text = f.read()
3 text

```

'챗봇\n\n문서\n토론\n읽기\n편집\n역사 보기\n위키백과, 우리 모두의 백과사전.\n\n\n가상 비서 챗봇\n챗봇(chatbot) 혹은 채터봇(chatterbot)은 음성이나 문자를 통한 인간과의 대화를 통해서 특정한 작업을 수행하도록 제작된 컴퓨터 프로그램이다. 토크봇(talkbot), 채터박스(chatbox) 혹은 그냥 봇(bot)라고도 한다.\n\n인간이 사용하는 언어를 이해해서 대화를 진행하는 자연언어처리 기술을 심각하게 적용하는 언어이해 방식, 입력받은 말에서 특정 단어나 어구를 검출하여 그에 맞는 미리 준비된 응답을 출력하는 검색 방식,\n\n그리고 각본을 미리 만들고 각본에 따라서 사용자의 입력에 대한 동작과 각본에 있는 응답을 출력하는 각본 방식이 있다.\n\n네이버 클라우드 플랫폼 카카오 토니 TME 스타 페이스브 메시저 스내키보 등에서 채보으 그도

```

1 mask = np.array(Image.open('./cloudMask.png'))
2 mask

```

```

array([[255, 255, 255,  0],
       [255, 255, 255,  0],
       [255, 255, 255,  0],
       ...,
       [255, 255, 255,  0],
       [255, 255, 255,  0],
       [255, 255, 255,  0]],

      [[255, 255, 255,  0],
       [255, 255, 255,  0],
       [255, 255, 255,  0],
       ...,
       [255, 255, 255,  0],
       [255, 255, 255,  0],
       [255, 255, 255,  0]],

```

```

[[255, 255, 255, 0],
 [255, 255, 255, 0],
 [255, 255, 255, 0],
 ...,
 [255, 255, 255, 0],
 [255, 255, 255, 0],
 [255, 255, 255, 0]],

...,

[[255, 255, 255, 0],
 [255, 255, 255, 0],
 [255, 255, 255, 0],
 ...,
 [255, 255, 255, 0],
 [255, 255, 255, 0],
 [255, 255, 255, 0]],

[[255, 255, 255, 0],
 [255, 255, 255, 0],
 [255, 255, 255, 0],
 ...,
 [255, 255, 255, 0],
 [255, 255, 255, 0],
 [255, 255, 255, 0]],

[[255, 255, 255, 0],
 [255, 255, 255, 0],
 [255, 255, 255, 0],
 ...,
 [255, 255, 255, 0],
 [255, 255, 255, 0],
 [255, 255, 255, 0]]], dtype=uint8)

```

```

1 from konlpy.tag import Okt
2 tagger = Okt()
3 nouns = tagger.nouns(text)
4 nounToStr = ' '.join(nouns)

```

```

1 from matplotlib.cm import ma
2 path = '/usr/share/fonts/truetype/nanum/NanumGothic.ttf'
3 s_words = STOPWORDS.union({
4     '이', '등', '의', '를', '것', '그', '이후', '도', '후', '모두', '또한', '따라서', '통한'
5 })
6 wordcloud = WordCloud(
7     font_path=path,
8     stopwords=s_words,
9     mask=mask,
10    background_color='white'
11 ).generate(nounToStr)

```

```
1 plt.imshow(wordcloud)
```

```
2 plt.axis('off')  
3 plt.show()
```



Colab 유료 제품 - [여기에서 계약 취소](#)

