

Fani Widodo

Lifelong Learner

FOLLOW:

NEXT STORY

Hacktoberfest itu ...



PREVIOUS STORY

Web Scrapping Dengan Python



ARCHIVES

November 2022

September 2021

August 2021

March 2021

February 2021

July 2020

RECENT POSTS

1 #JuaraGCP Season 8 (2022) lagi ...

1 #JuaraGCP Challenge 2021 itu ...

1 Twitter API

1 #JuaraGCP Crash Course itu ...

1 Melakukan Visualisasi Data Kasus COVID19 di Indonesia

PYTHON / TEKNOLOGI INFORMASI

Membuat Wordcloud Dengan Python

BY WIDODO · PUBLISHED MARCH 1, 2021 · UPDATED MARCH 5, 2021

Dalam tutorial kali ini, saya akan mencoba untuk membuat sebuah wordcloud menggunakan Python menggunakan library *wordcloud*.

Apa itu wordcloud? Berikut adalah definisinya menurut kamus cambridge:

“An electronic image that shows words used in a particular piece of electronic text or series of texts. The words are different sizes according to how often they are used in the text.”

Sederhananya, wordcloud adalah gambar yang menunjukkan daftar kata-kata yang digunakan dalam sebuah teks, umumnya semakin banyak kata yang digunakan semakin besar ukuran kata tersebut dalam gambar.

Dalam contoh ini saya akan menggunakan data hasil scrapping dari halaman website FEB Unpad dengan alamat <https://feb.unpad.ac.id/menristek-dorong-unpad-membangun-sdm-unggul-melalui-penciptaan-ekosistem-inovasi/>.

Data hasil scrapping sendiri sudah dibersihkan secara sederhana untuk menghilangkan tanda baca, mengecilkan huruf, dll. Teknik membersihkan teks hasil *scrapping* tidak akan dibahas disini karena selain cukup panjang juga tutorial kali ini hanya difokuskan pada pembuatan wordcloud menggunakan python.

Data yang digunakan dapat diunduh di [sini](#). Kode yang digunakan bisa diakses di [sini](#).

Mari Kita Mulai

Install library wordcloud

Langkah awal adalah kita melakukan instalasi library *wordcloud* dengan cara mengetikkan perintah berikut:

```
pip install wordcloud
```

Fungsi *wordcloud* membutuhkan masukan berupa text yang nantinya akan dihitung masing-masing katanya, yang selanjutnya akan digunakan untuk menggenerate hasil akhir berupa obyek.

Impor library yg digunakan

Sebelum mulai menggenerate wordcloud, kita perlu meng-impor beberapa library yang digunakan dalam program, ketikkan kode berikut:

```
from wordcloud import WordCloud
import matplotlib.pyplot as plt
```

Bisa dilihat diatas, selain kita memanggil library *wordcloud*, kita juga memanggil library *matplotlib*. Library *matplotlib* ini juga dipanggil karena hasil dari fungsi *wordcloud* nantinya digambar pada kanvas *matplotlib*.

Buka data text

Setelah mengimpor beberapa library yang nantinya akan digunakan, selanjutnya kita coba buka data teks yang akan digunakan untuk membuat wordcloud. Ketikkan kode dibawah ini untuk membuka file dan menyimpan isinya kedalam variabel **isi_text**.

MORE

KATEGORI

Big Data (1)

Cloud (4)

Coding (1)

Events (6)

Machine Learning (1)

NLP (1)

Opensource (1)

Python (4)

Teknologi Informasi (8)

TAGS

API Application Programming Interface big data Cloud coding Internet JuaraGCP machine learning nlp opensource python teknologi informasi twitter visualisasi data

```
f = open("berita_feb_unpad.txt", "r")

isi_text = f.read()
```

Untuk melihat apakah variabel **isi_text** sudah terisi, bisa dilakukan dengan mencoba mencetak variabel tersebut ke layar:

```
print(isi_text)
```

Maka akan didapatkan hasil sbb:

menristek dorong unpad membangun adm unggul melalui penciptaan ekosistem inovasi program doktor ilmu manajemen fakultas eko
nomi dan bisnis universitas padjadjaran menggelar kuliah perdana virtual bagi mahasiswa baru tahun akademik dengan tema aks
elerasi penguatan sistem inovasi nasional peluang dan tantangan pasca pandemi jumat prof bangbang menteri riset dan teknolo
gi ri atau menristek prof bangbang brodjonegoro saat menjadi pembicara kunci dalam acara kuliah perdana program doktor ilmu
manajemen fakultas ekonomi dan bisnis universitas padjadjaran secara virtual jumat menteri riset dan teknologi ri atau men
ristek prof bangbang brodjonegoro memaparkan strategi untuk mendorong indonesia keluar dari jurang negara pendapatan menenga
h menuju negara maju pada 2045 mendatang menurutnya perlu strategi khusus untuk mendukung pertumbuhan indonesia satu di ant
aranya adalah mengubah orientasi ekonomi yang semula berbasis sumber daya alam menjadi negara dengan perekonomian berbasis
inovasi inovasi sendiri harus dikembangkan ekosistem yang mendorong inovasi harus dibentuk dijaga dan disempurnakan untkap
prof bangbang saat menjadi pembicara kunci dalam acara kuliah perdana program doktor ilmu manajemen fakultas ekonomi dan bis
nis universitas padjadjaran secara virtual jumat meski demikian inovasi di indonesia belum banyak berkembang karena belum
adanya kesepahaman antara dunia penelitian dan dunia usaha karena itu penguatan inovasi perlu menekankan pada strategi kola
borasi triple helix atau kolaborasi yang melibatkan tiga unsur pemangku kepentingan yaitu akademisi dunia usaha dan pemerin
tah kolaborasi ini kata prof bangbang akan mendorong terciptanya kesepahaman antara aktivitas penelitian dan kebutuhan di pa
sar melalui fasilitasi penuh pemerintah tanpa kolaborasi ini dikhawatirkan sangat sedikit produk inovasi yang bisa diterima
oleh pasar lebih lanjut prof bangbang memaparkan inovasi indonesia perlu mempertimbangkan sejumlah hal pertama inovasi yang
dilahirkan merupakan teknologi tepat guna yang mudah dipakai murah serta relevan dengan kondisi setempat selanjutnya inovas
i bisa menggantikan peran impor mampu menciptakan nilai tambah dan komersialisasi beragam keanekaragaman hayati di indonesia
serta mengadopsi teknologi terkini dalam upaya mewujudkan strategi tersebut perguruan tinggi menjadi sektor yang terus dido
rong untuk menciptakan ekosistem inovasi prof bangbang menjelaskan pendidikan menjadi kunci untuk melahirkan sumber daya man
usia berkualitas yang akan menjadi penentu keberhasilan penguatan inovasi perguruan tinggi didorong untuk lebih sering laku
kan riset dan interaksi dengan dunia luar kata prof bangbang kuliah perdana bertajuk akselerasi penguatan sistem inovasi nas
ional peluang dan tantangan pasca-pandemi ini dibuka secara resmi oleh rektor unpad prof rina indiasutrisn acara juga diisi
dengan penyampaian materi kuliah oleh guru besar feb unpad yang juga atase pendidikan dan kebudayaan kbrri washington do ame
rika serikat prof popy rufaidah phd

Buat Wordcloud

Buat wordcloud dengan teks

Setelah variabel yang akan digunakan sebagai masukan fungsi *wordcloud* siap, maka kini kita siap untuk membuat obyek *wordcloud*. Silahkan ketik kode berikut untuk membuatnya:

```
wordcloud = WordCloud(width=1600, height=800, max_font_size=200,  
background_color='white')
```

Dalam fungsi tersebut terdapat beberapa argumen diantaranya *width*, *height*, *max_font_size*, dan *background_color*. Sesuai dengan namanya *width* dan *height* digunakan untuk mengatur lebar dan tinggi gambar wordcloud yang akan dihasilkan, kemudian *max_font_size* merupakan ukuran maksimal font dan *background_color* adalah untuk mengatur backgroud wordcloud yang nantinya dihasilkan.

Setelah obyek *wordcloud* obyek dibuat, kita selanjutnya panggil metode *generate()* untuk mengkalkulasi wordcloud dari argumen variabel *isi_text*.

```
wordcloud.generate(isi_text)
```

Menggambar hasil

Menampilkan worcloud kedalam gambar

Akhirnya kita siap untuk menampilkan hasilnya, disinilah library *matplotlib* akan digunakan. Kita dapat menggambar hasil wordcloud tersebut menggunakan fungsi *imshow()* yang disediakan oleh library *matplotlib*. Ketikkan kode berikut:

```
plt.figure(figsize=(12,10))

plt.imshow(wordcloud, interpolation='bilinear')

plt.axis("off")

plt.show()
```

Berikut adalah hasilnya:



Dari hasil diatas dapat dilihat bahwasanya kata **inovasi** dan **dan** menjadi kata yang paling banyak ditulis dalam berita. Karena proses pembersihan teks yang dilakukan masih sederhana, dalam wordcloud diatas masih banyak bertebaran kata yang tidak memiliki makna seperti **dan**, **di**, **yang**, **atau**, dll, yang memang sebaiknya dibersihkan dahulu dari teks sebelum kita membuat wordcloud.


Kesimpulan

Dengan menggunakan library *wordcloud* kita dapat dengan mudah membuat teks wordcloud di python dengan beberapa langkah sederhana:

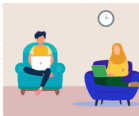
- Pertama adalah mempersiapkan dokumen teks yang akan digunakan untuk membuat wordcloud,
- Kedua adalah membuat obyek *wordcloud* dengan memberikan parameter dasar sebagai argumen, yang selanjutnya memanggil metode *generate()* untuk membuat obyek tersebut,
- Terakhir adalah menggunakan bantuan library *matplotlib* untuk menampilkan hasilnya.

Tags: python

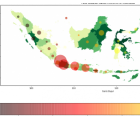
YOU MAY ALSO LIKE...



#JuaraGCP Challenge 2021 itu ...
SEPTEMBER 15, 2021



Melakukan Visualisasi Data Kasus COVID19 di Indonesia
MARCH 15, 2021



#JuaraGCP Season 8 (2022) lagi ...
NOVEMBER 24, 2022

