

Webcrawling and Wordcloud

April 22, 2021

```
[177]: from bs4 import BeautifulSoup
import urllib.request

title_list = []
url = 'https://www.google.com/search?
↳q=Coronavirus&rlz=1C10NGR_deDE948DE948&tbm=nws&start=0&sa=N&biw=631&bih=476&dpr=1.
↳5'
req = urllib.request.Request(url, headers ={'User-Agent': 'Mozilla/5.0 (X11;
↳Linux x86_64) AppleWebKit/537.11 (KHTML, like Gecko) Chrome/23.0.1271.64
↳Safari/537.11',
      'Accept': 'text/html,application/xhtml+xml,application/xml;q=0.9,*/*;q=0.
↳8',
      'Accept-Charset': 'ISO-8859-1,utf-8;q=0.7,*;q=0.3',
      'Accept-Encoding': 'none',
      'Accept-Language': 'en-US,en;q=0.8',
      'Connection': 'keep-alive'})
html = urllib.request.urlopen(req).read()
soup = BeautifulSoup(html, 'html.parser')
titles = soup.find_all(class_ = 'BNeawe vvjwJb AP7Wnd')
for title in titles:
    title_list.append(title.text)

#unerwünschte Data gelöscht
for j in range(11):
    title_list.remove(title_list[2])

num = [10,20,30,40,50,60,70,80,90,100,110,120,130,140,150,160,170,180,190,200]
for i in range(20) :
    url = 'https://www.google.com/search?
↳q=Coronavirus&rlz=1C10NGR_deDE948DE948&tbm=nws&start='+str(num[i])+ '&sa=N&biw=631&bih=476&d
↳5'
    req = urllib.request.Request(url, headers ={'User-Agent': 'Mozilla/5.0 (X11;
↳Linux x86_64) AppleWebKit/537.11 (KHTML, like Gecko) Chrome/23.0.1271.64
↳Safari/537.11',
```

```

    'Accept': 'text/html,application/xhtml+xml,application/xml;q=0.9,*/*;q=0.
↪8',
    'Accept-Charset': 'ISO-8859-1,utf-8;q=0.7,*;q=0.3',
    'Accept-Encoding': 'none',
    'Accept-Language': 'en-US,en;q=0.8',
    'Connection': 'keep-alive'})
html = urllib.request.urlopen(req).read()
soup = BeautifulSoup(html, 'html.parser')
titles = soup.find_all(class_ = 'BNeawe vvjwJb AP7Wnd')
for title in titles:
    title_list.append(title.text)

for j in range(len(title_list)):
    print(str(j)+' '+ title_list[j])
    print(' ')

#rObn4c rQMqOd
#BNeawe vvjwJb AP7Wnd

```

- 0 Neues Infektionsschutzgesetz: Mehrheit im Bundestag für die Notbremse
- 1 Härtere Corona-Maßnahmen in Bochum ab Donnerstag
- 2 Funktionsweise der Corona-Impfstoffe: So schützt die Impfung vor dem Coronavirus
- 3 Corona aktuell: RKI meldet 11 437 Neuinfektionen
- 4 Coronavirus-Pandemie: ++ Berlin will sich im Bundesrat enthalten ++ | tagesschau.de
- 5 Coronavirus - Bundesrat berät über Neufassung des Infektionsschutzgesetzes
- 6 Das Wichtigste im Liveblog: Wie die Welt gegen das Coronavirus kämpft
- 7 Coronavirus-Blog: Die Lage am Mittwoch, 21. April | NDR.de - Nachrichten
- 8 Coronavirus: So viele freie Intensivbetten wie noch nie seit Pandemie-Beginn
- 9 Coronavirus: „Da ist noch Luft nach oben“: Biontech rüstet sich für weiteren Ausbau der Impfstoff-Produktion
- 10 Coronavirus: „Da ist noch Luft nach oben“: Biontech rüstet sich für weiteren Ausbau der Impfstoff-Produktion
- 11 Corona-Zahlen NRW: Inzidenzwert weiter hoch - Dienstag bei 168,4

- 12 Corona-Update: Sechs weitere Todesfälle im Kreis, 3 Neuinfektionen in Haltern
- 13 Dortmunderin kritisiert Gesundheitsamt: „Dem Virus spielt das in die Karten“
- 14 Risiko von Nebenwirkungen: Ciesek hält Gefahr durch Impfung für überschätzt
- 15 Intensivbetten statt Inzidenz: Was sagt mehr über die Pandemielage aus?
- 16 Bochum: Migranten und Corona - die ersten Reaktionen
- 17 Funktionsweise der Corona-Impfstoffe: So schützt die Impfung vor dem Coronavirus
- 18 Coronavirus in Nahost: Viel Skepsis gegen Impfungen
- 19 7-Tage-Inzidenz: Epidemiologe zweifelt an Orientierungswert in derzeitiger Corona-Lage
- 20 Coronavirus-Grafiken: Wo Deutschland Rot ist
- 21 Coronavirus: Gesundheitsamt meldet zwei weitere Todesfälle
- 22 Coronavirus in Europa - Frankreich will Corona-Beschränkungen ab Mai schrittweise lockern
- 23 Gene in Asien - Coronavirus-Epidemie schon vor 25.000 Jahren?
- 24 Covid-19 - Welche Rolle Aerosole bei der Übertragung des Coronavirus spielen
- 25 Coronavirus - Bundestag beschließt Änderung des Infektionsschutz-Gesetzes
- 26 Coronavirus erreicht Mount Everest
- 27 Coronavirus: Katastrophale Impfquote in Hessen - Wieso wird so langsam geimpft?
- 28 Völlig neuer Ansatz: So wollen Forscher aus Bayreuth das Coronavirus zerstören
- 29 MV gibt Astrazeneca ab sofort für alle Altersgruppen frei
- 30 Coronavirus Baden-Württemberg: Inzidenz steuert auf 200er-Marke zu
- 31 Coronavirus: Notbremse greift - Folgen für Schulen unklar
- 32 Völlig neuer Ansatz: So wollen Forscher aus Bayreuth das Coronavirus zerstören

- 33 Coronavirus: Diese Gruppen stecken sich an - extrem unterschiedliche München-Verteilung
- 34 Lübecks Uni-Forscher Hilgenfeld legt Besonderheit des Coronavirus offen
- 35 Wie gefährlich ist das Coronavirus für Kinder?
- 36 Deutschlandkarte zeigt, wie unterschiedlich sich das Coronavirus in den Altersgruppen verbreitet
- 37 Corona-Quarantäne in NRW: Sonderregeln für Geimpfte - Das müssen Sie wissen
- 38 Coronavirus-Bekämpfung: Was Gurgeln wirklich bringt
- 39 Coronavirus: Erneut mehr Neuinfektionen als in der Vorwoche
- 40 Kampf gegen das Coronavirus - J&J-Impfstoff soll bald in Deutschland eingesetzt werden
- 41 Corona aktuell: Eilantrag gegen Aufbaufonds abgewiesen
- 42 Coronavirus: Robert Koch-Institut meldet wieder weniger Neuinfektionen
- 43 Coronavirus in Deutschland: NRW meldet unvollständige Corona ...
- 44 Coronavirus: 132 Neuinfektionen im Kreis Borken
- 45 Coronavirus - Bei Schnelltests ist Geduld gefragt
- 46 Liveblog zum Coronavirus in Castrop-Rauxel: Die Lage im Überblick
- 47 Das Corona-Update vom 21. April
- 48 Dortmund: Coronavirus trotz Impfung?
- 49 Putin ruft zu Impfungen gegen Coronavirus auf
- 50 Oberhavel: Coronavirus: 92-Jährige aus Oranienburg stirbt - 21 Neuinfektionen in Oberhavel
- 51 Coronavirus: Wie gefährlich ist die Mutation B.1.617?
- 52 Coronavirus: Ausgangssperren nicht effektiv? Neue Studienergebnisse lassen aufhorchen
- 53 Coronavirus: BRK richtet Teststation in Neustadt/WN ein

- 54 Coronavirus Landkreis Leipzig: 13.065 bestätigte Fälle (Stand 21 ...
- 55 Funktionsweise der Corona-Impfstoffe: So schützt die Impfung vor ...
- 56 Coronavirus-Pandemie: ++ Niederlande wollen lockern ++ ...
- 57 Coronavirus: Mobiler Impf-Registrierungsservice ist seit einer ...
- 58 21.4.2021 Coronavirus: 21 Neuinfektionen, darunter 9 in Sinzig, 1 in ...
- 59 Robert Koch-Institut - Bereits knapp 95 Prozent Anteil der britischen Coronavirus-Variante
- 60 Coronavirus in Bayern: Kabinett lockert Regeln nicht
- 61 Coronavirus-Liveticker:+++ 01:38 Indien: 22 Covid-Patienten sterben wegen Leck in Sauerstofftank +++
- 62 Coronavirus-Blog: Die Lage am Sonnabend, 17. April | NDR.de ...
- 63 Daniela Katzenberger, Coronavirus und Schönheits-OPs: Neue Folgen ihrer Doku-Soap
- 64 Coronavirus-Pandemie: ++Türkei meldet Höchstwert bei ...
- 65 Coronavirus - Bundesrat berät über Neufassung des Infektionsschutzgesetzes
- 66 Coronavirus: Impfungen für Menschen über 60 Jahren starten
- 67 Besorgniserregende Mutationen: Das Verwandlungsrepertoire des ...
- 68 Schwesig: Coronavirus ist Staatsfeind Nummer 1
- 69 Coronavirus in Stuttgart und Region - Viele Zimmer in den Studentenwohnheimen stehen leer
- 70 Coronavirus in Baden-Württemberg - Inzidenz leicht gesunken - 12 Kreise über 200
- 71 Interview mit Ernährungsspezialist: Schützen Vitamin-D-Kapseln vor Corona?
- 72 Coronavirus in der Schweiz - In diesen drei Phasen will der ...
- 73 Warum sich das Coronavirus verändert: Mutieren gehört zum Geschäft
- 74 Modellregionen Eckernförde und Schlei starten Tourismusprojekt
- 75 Corona-Ausbruch in Klinik in Dortmund: Besucher trägt Coronavirus mit

falsch-negativem Schnelltest in Klinikum

76 Corona-Pandemie - Jeder Fünfte in Deutschland mindestens einfach gegen Coronavirus geimpft

77 Coronavirus-Pandemie: ++ Hälfte der Erwachsenen in den USA ...

78 Coronavirus: Modellprojekten droht der Lockdown | tagesschau.de

79 Alle Infos zum Coronavirus im Kreis Görlitz

80 Coronavirus: Über 50 Infektionsfälle in Bad Honnef

81 Ansturm bei den Schnelltests an der Müritz bleibt aus

82 Corona im Kreis Calw: Mann stirbt an Infektion mit Coronavirus

83 Coronavirus in Dresden: Abiturprüfungen in Sachsen unter Pandemie-Bedingungen stehen an

84 Coronavirus-Bekämpfung: Was Gurgeln wirklich bringt

85 Coronavirus-Blog: Die Lage am Mittwoch, 14. April | NDR.de ...

86 Verordnung zur Eindämmung der Ausbreitung des Coronavirus ...

87 Coronavirus: Ein weiterer Todesfall, Inzidenz nur noch knapp unter ...

88 Corona-Tote und Neuinfektionen: Castrop-Rauxel bleibt Hotspot im Kreis

89 RKI-Präsident Wieler: Coronavirus "lässt sich nicht ausrotten"

90 Coronavirus-Fallzahlen: Inzidenz im Landkreis Südwestpfalz am 21 ...

91 Coronavirus-Pandemie: ++ Notbremse nicht scharf genug? ++ ...

92 Coronavirus: Was steckt hinter Covid-19? Alle Infos zu Übertragung, Symptomen und Bekämpfung

93 Coronavirus: Sind Sie schon geimpft?

94 Einheitliche Coronaregeln: Bundesnotbremse versagt, bevor sie überhaupt gezogen wurde

95 Coronavirus in Deutschland: Sieben-Tage-Inzidenz steigt auf 140,9 ...

96 Reisewarnungen - RKI aktualisiert Liste der Coronavirus-Risikogebiete

- 97 Dritte Infektionswelle in Deutschland: Das Impf-Paradoxon
- 98 Coronavirus: Mehr Infizierte und Tote unter finanziell Benachteiligten
- 99 Tempo zieht weiter an: Über 50.000 Impfungen gegen das Coronavirus in Hamm
- 100 (84) Coronavirus-Update: Nicht auf Tests und Impfungen verlassen
- 101 Kreisverwaltung Warendorf: Coronavirus: Weiter steigende Fallzahlen
- 102 Coronavirus in Deutschland: Die dritte Welle ist nicht gebrochen
- 103 Strategie gegen Corona: Schnelltests sind unzuverlässig - aber was sind die Alternativen?
- 104 Coronavirus-Blog: Die Lage am Sonntag, 11. April | NDR.de ...
- 105 Coronavirus in Deutschland und der Welt: Indien verzeichnet fast ...
- 106 Coronavirus-Pandemie: ++ Schwangere sollen mRNA-Impfung ...
- 107 Notbremse: Ab Donnerstag gelten diese Regeln » Coronavirus
- 108 Coronavirus: Keine Übertragung über Trinkwasser
- 109 Coronavirus-Pandemie: ++ Stiko hält an Impfpriorisierung fest ++ ...
- 110 Covid-19-Impfstoff Moderna: Impfstoffvarianten neutralisieren ...
- 111 Coronavirus-Blog: Die Lage am Sonnabend, 10. April | NDR.de ...
- 112 Coronavirus-Schutz durch Sport? Lauterbach stellt brisante Studie vor - „klar belegt“
- 113 Science-Fiction statt Test: Pentagon-Chip erkennt Coronavirus
- 114 (83) Coronavirus-Update: Impfeffekt noch nicht in Sicht | NDR.de ...
- 115 Coronavirus-News: Die zentrale Corona-Gedenkfeier im Livestream
- 116 Coronapandemie: Präsident der Intensivmediziner warnt vor »dramatischer Lage« in Krankenhäusern
- 117 Coronavirus-Ticker am Mittwoch: 50 Millionen zusätzliche Biontech-Dosen in diesem Quartal
- 118 Corona aktuell: RKI meldet 11 437 Neuinfektionen

- 119 Coronavirus-Blog: Die Lage am Ostermontag, 5. April | NDR.de ...
- 120 Schleswig-Holsteins Datenschutzbeauftragte kritisiert Luca-App
- 121 Coronavirus-Pandemie: ++ Baden-Württemberg will Regeln für ...
- 122 (82) Coronavirus-Update: Die Lage ist ernst | NDR.de - Nachrichten
- 123 Coronavirus-Pandemie: ++ Lauterbach mit AstraZeneca geimpft ++ ...
- 124 Coronavirus in Baden-Württemberg - Land beschließt Ausnahmen für vollständig Geimpfte
- 125 Coronavirus: Experten bezeichnen Studie zu Asthmamedikamenten als vielversprechend
- 126 Coronavirus Baden-Württemberg: Wackelt Demo-Verbot in Stuttgart? Zahlen explodieren
- 127 Merkel mit Astra-Zeneca gegen Coronavirus geimpft
- 128 Coronavirus-Pandemie: ++ Müller sieht Ausgangssperre kritisch ++ ...
- 129 Coronavirus in Bayern: Schulen zu ab Inzidenz 100 oder 200? Verwirrung bei der neuen Notbremse
- 130 Kampf gegen das Coronavirus - Stuttgart erlässt eine Ausgangssperre
- 131 Coronavirus im Kreis Recklinghausen: Inzidenz im Kreis knapp unter 200 - Ausgangsbeschränkungen erst ab 22 Uhr
- 132 Coronavirus in Stuttgart - Inzidenz springt auf bisher höchsten Wert
- 133 Traunstein: Wie gefährlich ist das Coronavirus für Kinder? Das sind die Fakten
- 134 RKI meldet 11 437 Neuinfektionen
- 135 Coronavirus: Mutationen immer häufiger in Baden-Württemberg
- 136 TV-Moderator: Günther Jauch mit Coronavirus infiziert
- 137 Coronavirus in Deutschland: Bayern will ebenfalls AstraZeneca-Impfpriorisierung aufheben
- 138 Coronavirus weltweit: Frankreich plant trotz hoher Inzidenz Lockerungen im Mai

- 139 Coronavirus im Kreis Recklinghausen: Schon 15 Kommunen in Nordrhein-Westfalen liegen bei der Inzidenz über 200
- 140 Coronavirus-Pandemie: ++ Papst ruft zu Hoffnung auf ++ ...
- 141 Corona in der Region und Baden-Württemberg (Stand: 8. April ...
- 142 Streit um Corona-Vakzine: Österreich droht EU mit Impfstoffblockade
- 143 Coronavirus im Landkreis Weilheim-Schongau - alle Informationen im News-Ticker
- 144 Coronavirus-Pandemie: ++ Alle New Yorker impfberechtigt ++ ...
- 145 Coronavirus in der Region: Inzidenz in Berlin jetzt bei 152 - in ...
- 146 Coronavirus-Pandemie: ++ Pilotprojekt in Tübingen läuft weiter ++ ...
- 147 Coronavirus in Baden-Württemberg - Wie kommt es zu diesen Infektionszahlen?
- 148 Corona Bayern: Intensivbetten in Nürnberg werden knapp
- 149 Coronavirus im Landkreis Bad Tölz-Wolfratshausen: Neuer Todesfall - Inzidenz nähert sich 100
- 150 Beerdigung am Dienstag: Weiteres Mitglied der Großfamilie stirbt an ...
- 151 Coronavirus-Pandemie: ++ Hans verteidigt Lockerungspläne ++ ...
- 152 Newsblog zum Coronavirus +++ Robert-Koch-Institut meldet 9.609 ...
- 153 Corona: Inzidenz steigt sprunghaft an - MK wieder Hotspot Nummer eins in NRW
- 154 Corona: Coronavirus trifft jetzt die jüngeren Remscheid - In keiner NRW-Stadt verbreitet sich das Virus sch...
- 155 Corona Bayern: Verfassungsbeschwerde der Freien Wähler
- 156 Inzidenz in Deutschland bei 136,4
- 157 Coronavirus: WHO fordert Verkaufsverbot von Wildtieren
- 158 Corona: Landesregierung konkretisiert Notbremse
- 159 Coronavirus in Deutschland: RKI meldet 23.804 Neuinfektionen ...
- 160 Coronavirus im Kreis Recklinghausen: Mehr aktuelle Fälle als Weihnachten -

171 Neuinfektionen binnen 24 Stu...

161 Liveblog: ++ STIKO rät Jüngeren von AstraZeneca als zweite Dosis ...

162 Coronavirus in Stuttgart und Region - Wegen Astrazeneca: Terminchaos in Impfzentren

163 Corona Bayern: Bis zu 50 000 Impfdosen für Betriebe

164 Kampf gegen Coronavirus: Wie läuft es mit den Impfungen ...

165 Kreisverwaltung Warendorf: Coronavirus: Gesundheitsamt meldet ...

166 Coronavirus-Liveticker:+++ 07:39 Berliner Verfassungsschutz beobachtet Protestbewegung +++

167 Hamburg: Schulbehörde hält Schulschließungen für möglich

168 463 neue Fälle in Berlin - ein Bezirk bei Inzidenz über 200

169 20:47 Gericht: Keine Ausnahme von Corona-Regeln für Geimpfte

170 Günther Jauch: Er hat sich mit dem Coronavirus infiziert

171 RKI registriert 8497 Corona-Neuinfektionen

172 Coronavirus-Liveticker:+++ 18:43 Lehrer loben niedrigen Grenzwert für Schulschließungen +++

173 Empörung über illegale Luxusdinner in Paris

174 (79) Coronavirus-Update: Angriffspunkte für das Virus | NDR.de ...

175 Coronavirus-Blog: Die Lage am Donnerstag, 21. Januar | NDR.de ...

176 Hamburger Forscher: Coronavirus stammt wohl aus Labor

177 Coronavirus: Schnelltest-Pläne vorerst gestoppt | tagesschau.de

178 Coronavirus-Blog: Die Lage am Freitag, 1. Januar | NDR.de ...

179 Coronavirus-Blog: Die Lage am Donnerstag, 7. Januar | NDR.de ...

180 Coronavirus-Mutation: Fragen und Antworten

181 Coronavirus-Blog: Die Lage am Montag, 15. Februar | NDR.de ...

182 Corona-Impfung: Wer kann sich wann impfen lassen?

183 Minister Garg besorgt: SH wieder Risikogebiet

184 Coronavirus-Blog: Die Lage am Sonntag, 27. Dezember

185 Coronavirus: Was können Selbst-Schnelltests? | tagesschau.de

186 Coronavirus-Blog: Die Lage am Montag, 8. Februar | NDR.de ...

187 Coronavirus-Blog: Die Lage am Freitag, 25. Dezember | NDR.de ...

188 Corona-Impfung in SH: Wann auch über 60-Jährige dran sein könnten

189 Coronavirus: Symptome, Ansteckung, Ausbreitung, Mutation - alle Infos

190 Coronavirus-Blog: Die Lage am Dienstag, 23. Februar | NDR.de ...

191 Impfstart: Mit einem Pieks gegen das Coronavirus

192 Drostens im Corona-Podcast: AstraZeneca ist kein zweitrangiger ...

193 Coronavirus-Blog: Die Lage am Montag, 30. November | NDR.de ...

194 Corona-Fälle bei Airbus: Virus-Mutation nachgewiesen

195 Coronavirus-Blog: Die Lage am Donnerstag, 11. Februar | NDR.de ...

196 Vertane Chance: Coronavirus-Nachweis im Abwasser

197 Können Geimpfte das Coronavirus weiterverbreiten?

198 Coronavirus-Blog: Die Lage am Freitag, 5. Februar | NDR.de ...

199 Schutzimpfung gegen das Coronavirus SARS-CoV-2 ...

200 Coronavirus-Blog: Die Lage am Donnerstag, 25. Februar | NDR.de ...

201 Corona-Impfung: Die wichtigsten Fragen und Antworten

202 (70) Coronavirus-Update: Die Mutanten im Blick behalten | NDR.de ...

203 Südafrikanische Corona-Mutante in Hamburg nachgewiesen

204 303 neue Corona-Fälle in Hamburg

205 Vier Corona-Impfstoffe im Vergleich

206 Studiendaten aus Israel: Biontech-Impfstoff stoppt Virusübertragung zu 89,4

Prozent

207 Coronavirus: Verteilung in Hamburgs Bezirken

208 Coronavirus-Blog: Die Lage am Sonnabend, 9. Januar | NDR.de ...

209 Coronavirus-Mutanten: Warum die sinkenden Infiziertenzahlen trügerisch sind

```
[178]: f = open("coronaviursnewsheadline.txt", "w", encoding = 'utf8')
f.write(str(title_list))
f.close()
```

```
[179]: import numpy as np #numpy array
import os
import re
from PIL import Image
from os import path
from wordcloud import WordCloud
import matplotlib.pyplot as plt
from collections import Counter
```

```
[186]: mask = np.array(Image.open("hearts.png"))
```

```
[187]: f = open('coronaviursnewsheadline.txt', 'r', encoding = 'utf8')
txt = f.readline()
f.close()

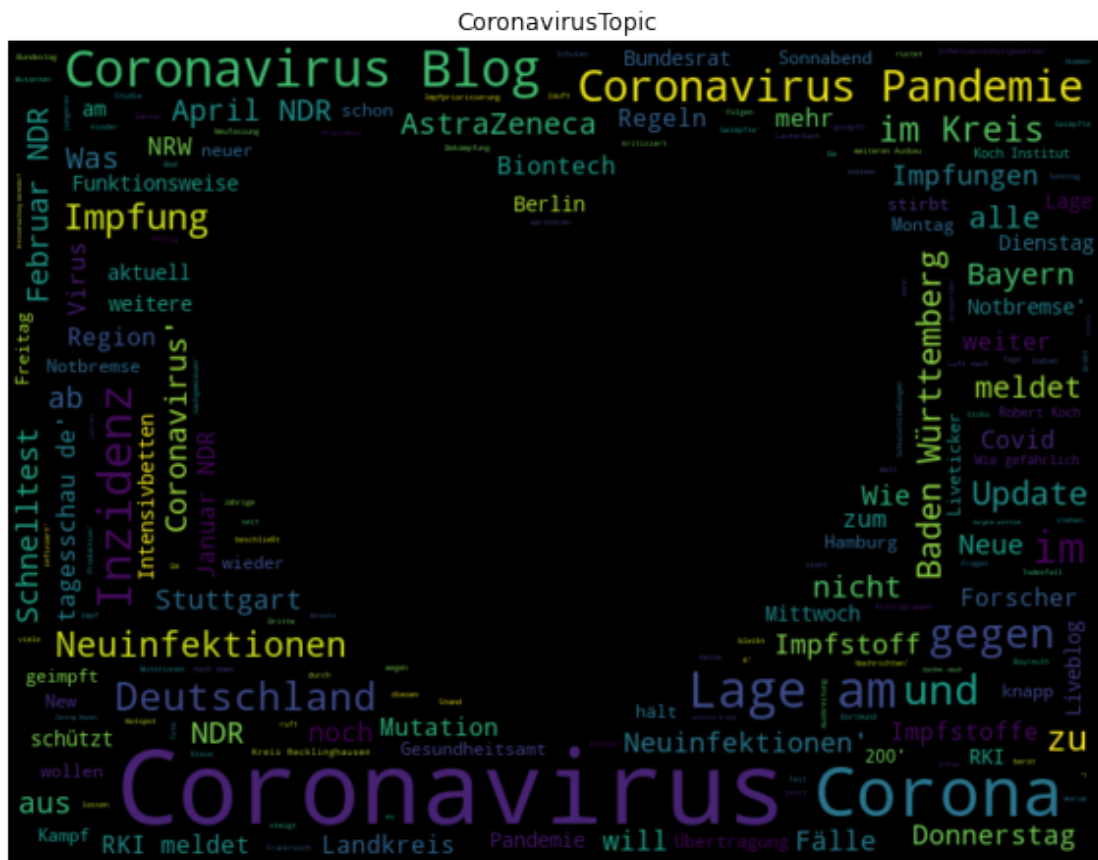
#UTF-8 (Abkürzung für 8-Bit UCS Transformation Format, wobei UCS wiederum
→Universal Coded Character Set abkürzt) ist die am weitesten verbreitete
→Kodierung für Unicode-Zeichen (Unicode und UCS sind praktisch identisch)
```

```
[201]: Counter(txt)
stopwords = ['[', ']', ':',
→',', '+', '-', 'der', 'die', 'das', 'dem', 'den', 'des', 'derren', 'mit', 'Sie', 'über', 'ein', 'ist', 'in',
```

```
[203]: wc = WordCloud(max_words = 1000, mask = mask, stopwords = stopwords, margin =
→10, random_state = 1, background_color = 'black', contour_color = 'black').
→generate(txt)

wc.to_file("coronavirustopic.png")
plt.axis('off')
plt.figure(figsize = (10,10))
plt.title("CoronavirusTopic")
plt.imshow(wc, interpolation = "bilinear")
plt.axis('off')
plt.show
```

```
[203]: <function matplotlib.pyplot.show(close=None, block=None)>
```



[]:

[]: