

July 13, 2021

SET DE TWEETS FRANCOPHONES RELATIF A LA CRISE DE COVID-19 A DES FINS DE RECHERCHE

Aurélie Montarnal; Julien Coche; Sandrine Bubendorff; Nathan Thubert; Marie-Lys Camozzi; Caroline Rizza

Contexte

Ce dataset est mis à disposition dans le cadre du projet MESCOV « les media sociaux lors de la crise Covid-19 » financé par le Comité analyse, recherche et expertise (CARE) du Ministère de l'Education Supérieure, de la Recherche et de l'Innovation. Le projet MESCOV traite des aspects création et circulation de l'information sur les media sociaux lors de la crise COVID-19, des initiatives citoyennes qui y ont émergé et des pratiques des professionnels de la gestion de crise associées (notamment Service d'Incendie et de Secours et Préfecture).

C'est un projet pluridisciplinaire qui mobilise à la fois les Sciences de l'Informatique et de la Donnée pour le module base de données et algorithme d'apprentissage automatique et les Sciences Humaines et Sociales pour la partie documentation des mécanismes de création, de circulation et de vérification de l'information sur les media sociaux, l'émergence d'initiatives citoyennes et l'utilisation des médias sociaux par les institutionnels (Camozzi, et al., 2020, à paraître).

Acquisition des données et constitution du jeu de tweets

Ce jeu de tweets a été généré à partir du dataset proposé par Banda et al. (2020) récolté en temps-réel. Ce dataset a été constitué à partir des mots-clés suivants : COVID19, CoronavirusPandemic, COVID-19, 2019nCoV, CoronaOutbreak, coronavirus, WuhanVirus, covid19, coronaviruspandemic, covid-19, 2019ncov, coronaoutbreak, wuhanvirus.

Les tweets utilisés sont ceux publiés entre le 22 mars et le 24 juin 2020.

Le jeu de données que nous proposons a été hydraté en utilisant l'outil Twarc, et seuls les tweets en langue française ont été conservés. Il comprend 2.950.157 tweets au 15 juillet 2020, date de sa création.

Conditions d'utilisation et citation du set de tweets

Si vous citez ou réutilisez ce dataset merci de mentionner Montarnal, A., Coche, J., Bubendorff, S. Thubert, N., Camozzi, M-L., & Rizza, C. (2020) « Set de tweets francophones relatif à la crise de covid-19 à des fins de recherche ».

Ce jeu de tweets a été créé sur la base légale de l'intérêt public. Il est fourni tel quel, suivant les règlements de Twitter.

Seuls les identifiants des tweets sont fournis au format csv. L'hydratation du dataset est possible, par exemple en utilisant le projet Python Twarc (<https://github.com/DocNow/twarc>).

Il ne contient pas de données personnelles mais en réhydratant ce dataset, les chercheurs pourront retrouver les auteurs des tweets à partir de l'ID du tweet associé. Ce jeu de tweets ne peut être utilisé qu'à des fins non-commerciales et de recherche. Les chercheurs seront responsables de leur traitement.

Références

Banda J-M., et al., "A large-scale COVID-19 Twitter chatter dataset for open scientific research -- an international collaboration," arXiv:2004.03688 [cs], Nov. 2020, Accessed: Nov. 24, 2020. [Online]. Available: <http://arxiv.org/abs/2004.03688>.

Camozzi M-L., Thubert N., Coche J., Bubendorff S., Montarnal A., & Rizza C. (2020) Les media sociaux lors de la crise sanitaire de Covid-19 : Circulation de l'information et initiatives citoyennes. i3 Working Papers Series, 16-SES-01

Projet MESCOV financé par le Comité analyse, recherche et expertise (CARE) du Ministère de l'Education Supérieure, de la Recherche et de l'Innovation.

90

views

12

downloads

[See more details...](#)

Indexed in

Publication date:

July 13, 2021

DOI:

DOI: [10.5281/zenodo.5094969](https://doi.org/10.5281/zenodo.5094969)

Keyword(s):

[covid19](#) [covid](#) [tweets](#) [twitter](#) [pandemic](#) [franch](#)

License (for files):

[Creative Commons Attribution 4.0 International](#)

Versions

Version 1 July 13, 2021

[10.5281/zenodo.5094969](https://doi.org/10.5281/zenodo.5094969)

Cite all versions? You can cite all versions by using the DOI [10.5281/zenodo.5094968](https://doi.org/10.5281/zenodo.5094968). This DOI represents all versions, and will always resolve to the latest one. Read more.

Share



Cite as

Aurélie Montarnal, Julien Coche, Sandrine Bubendorff, Nathan Thubert, Marie-Lys Camozzi, & Caroline Rizza. (2021). SET DE TWEETS FRANCOPHONES RELATIF A LA CRISE DE COVID-19 A DES FINS DE RECHERCHE [Data set]. Zenodo. <https://doi.org/10.5281/zenodo.5094969>

Export

[BibTeX](#) [CSL](#) [DataCite](#) [Dublin Core](#) [DCAT](#)
[JSON](#) [JSON-LD](#) [GeoJSON](#) [MARCXML](#)
[Mendeley](#)

Files (74.0 MB)

Name

Size

datasetCOVID.csv

73.8 MB

[Preview](#)
[Download](#)

md5:d59d29a63ee65ba7a4dc62b01cf88383

notice_datasetCOVID.pdf

265.5 kB

[Preview](#)
[Download](#)

md5:3204d653d2417b9ad3fdbd45b10a5541

[About](#)
[Blog](#)
[Help](#)
[Developers](#)
[Contribute](#)
[About](#)
[Blog](#)
[FAQ](#)
[REST API](#)
[GitHub](#)
[Policies](#)
[Features](#)
[OAI-PMH](#)
[Donate](#)
[Infrastructure](#)
[Support](#)
[Principles](#)
[Contact](#)

Funded by

