A Global COVID-19 Observatory, Monitoring the Pandemics Through Text Mining and Visualization

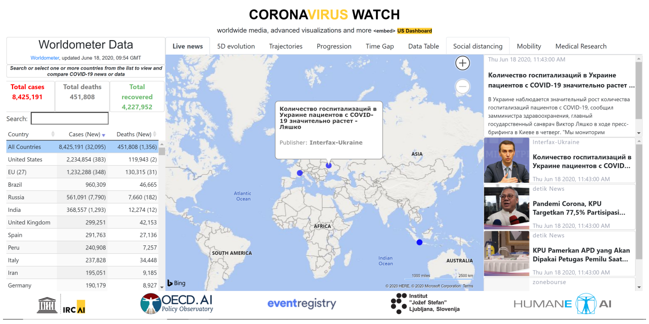
M. Besher Massri[1,2], Joao Pita Costa, Marko Grobelnik, Janez Brank, Luka Stopar and Andrej Bauer[3]
E-mail: besher.massri@ijs.si, andrej.bauer@andrej.com
[1]Jožef Stefan Institute, Slovenia
[2]Jožef Stefan International Postgraduate School, Ljubljana, Slovenia
[3]University of Ljubljana, Slovenia

Keywords: text mining, data analytics, data visualisation, public health, Coronavirus, COVID-19, epidemic intelligence

Received: November 23, 2020

The global health situation due to the SARS-COV-2 pandemic motivated an unprecedented contribution of science and technology from companies and communities all over the world to fight COVID-19. In this paper, we present the impactful role of text mining and data analytics, exposed publicly through IRCAI's Coronavirus Watch portal. We will discuss the available technology and methodology, as well as the ongoing research based on the collected data.

Povzetek: Opisana je vloga rudarjenja besedil in podatkovne analitike na primeru portala IRCAI's Coronavirus Watch.


CORONAVIRUS WATCH
worldwide media, advanced visualizations and more <embed> US Dashboard

Worldometer Data
Worldometer, updated June 18, 2020, 09:54 GMT
Search or select one or more countries from the list to view and compare COVID-19 news or data

Total cases	Total deaths	Total recovered
8,425,191	451,808	4,227,952

Search:

Country	Cases (New)	Deaths (New)
All Countries	8,425,191 (32,095)	451,808 (1,356)
United States	2,234,854 (383)	119,943 (2)
EU (27)	1,232,288 (348)	130,315 (31)
Brazil	960,309	46,665
Russia	561,091 (7,790)	7,660 (182)
India	368,557 (1,293)	12,274 (12)
United Kingdom	299,251	42,153
Spain	291,763	27,136
Peru	240,908	7,257
Italy	237,828	34,448
Iran	195,051	9,185
Germany	190,179	8,927

Live news | 5D evolution | Trajectories | Progression | Time Gap | Data Table | Social distancing | Mobility | Medical Research

Количество госпитализаций в Украине пациентов с COVID-19 значительно растет - Ляшко
Publisher: Interfax-Ukraine

Thu Jun 18 2020, 11:43:00 AM
Количество госпитализаций в Украине пациентов с COVID-19 значительно растет ...
В Украине наблюдается значительный рост количества госпитализаций пациентов с COVID-19, сообщил замминистра здравоохранения, главный государственный санврач Виктор Ляшко в ходе пресс-брифинга в Киеве в четверг. "Мы мониторим

Interfax-Ukraine
Количество госпитализаций в Украине пациентов с COVID...
Thu Jun 18 2020, 11:43:00 AM

detik News
Pandemi Corona, KPU Targetkan 77,5% Partisipasi...
Thu Jun 18 2020, 11:43:00 AM

detik News
KPU Pamerkan APD yang Akan Dipakai Petugas Pemilu Saat...
Thu Jun 18 2020, 11:43:00 AM

zonebourse

Atlantic Ocean
ASIA
AFRICA
SOUTH AMERICA
Indian Ocean
AUSTRALIA
Bing
1000 miles
2500 km
© 2020 HERE, © 2020 Microsoft Corporation Terms
IRCAI
OECD.AI Policy Observatory
eventregistry
Institut "Jožef Stefan" Ljubljana, Slovenija
HUMANE AI

Figure 1: Coronavirus Watch portal

1 Introduction

When the World Health Organization (WHO) announced the global COVID-19 pandemic on March 11th 2020 [36], following the rising incidence of the SARS-COV-2 in Europe, the world started reading and talking about the new Coronavirus. The arrival of the epidemic to Europe scaled out the news published about the topic, while public health institutions and governmental agencies had to look for existing reliable solutions that could help them plan their actions and the consequences of these. Technological companies and scientific communities invested efforts in making available tools (e.g. the GIS [2] later adopted by the World Health Organisation (WHO)), challenges (e.g. the Kaggle COVID-19 competition [22]), and scientific reports and data (e.g. the repositories medRxiv [24] and Zenodo [38]). In this paper we discuss the Coronavirus Watch portal [21], made available by the UNESCO AI Research Institute (IRCAI), comprehending several data exploration dashboards related to the SARS-COV-2 worldwide pandemic (see the main portal in Figure 1). This platform aims to expose the different perspectives on the data generated and trigger actions that can contribute to a better understanding of the behavior of the disease.

2 Related work

With with growing dimension of the public health problem caused by the SARS-COV-2, the global effort multiplied worldwide: from the publicly shared research in, e.g., Zenodo [38] medRxiv [24] and IEEE Xplore [20]; to a diversity of datasets made available by the global agencies (as e.g., the European Commission [3] and the ECDC [6]), the tech giants (as, e.g., Google [7] and Facebook [15]), institutional initiatives (as, e.g., the OxCOVID19 project [34]), and private initiatives acting globally (e.g. [37]) and locally (e.g., the slovenian initiative [32]). With similar aim to contribute for the global effort, many platforms have been made publicly available over the internet to monitor aspects of the COVID-19 pandemics are mostly focusing on data visualization based on the incidence of the disease and the death rate worldwide (e.g., the CoronaTracker [4]). The limitations of the available tools are potentially due to the lack of resolution of the data in aspects like the geographic location of reported cases, the commodities (i.e., other diseases that also influence the death of the patient), the frequency of the data, etc. On the other hand, it was not common to monitor the epidemic through the worldwide news (with some exceptions as, e.g., the Ravenpack Coronavirus News Monitor [31]). This is a rather important perspective mostly due to the great importance that the related misinformation - known as *infodemia*[1] - has been playing a role in the context of this pandemics.

The Coronavirus Watch portal suggests the association of reported incidence with worldwide published news per country, which allows for real-time analysis of the epidemic situation and its impact on public health (in which specific topics like mental health and diabetes are important related matters) but also in other domains (such as economy, social inequalities, etc.). This news monitoring is based on state-of-the-art text mining technology aligned with the validation of domain experts that ensures the relevance of the customized stream of collected news.

Moreover, the Coronavirus Watch portal offers the user other perspectives of the epidemic monitoring, such as the insights from the published biomedical research that will help the user to better understand the disease and its impact on other health conditions. While related work was promoted in [22] in relation with the COVID-19, and is offered in general by MEDLINE mining tools (e.g., MeSH Now [25]), there seems to be no dedicated tool to the monitoring and mining of COVID-19 - related research as that presented here.

3 Description of data

The nature of the proposed global observatory system entails a range of heterogeneous data sources that entangle to provide a perspective as complete as possible on the matters of the COVID-19 pandemics. In the following section we discuss these data sources and their potential contributions.

3.1 Historical COVID-19 data

To perform an analysis of the growth of the coronavirus, we use the historical data on the daily number of cases of infections and deaths. This data is retrieved from a GitHub repository made available by the John Hopkins University [5]. The data source is based mainly on the official data from the World Health Organization (WHO)[28] along with some other sources provided by, e.g., the Center for Disease and Control[17], and Worldometer[37], among others. This data provides the basis for all the system's functionality that depended on the statistical information about SARS-COV-2 incidence.

3.2 Live data from worldometer

Apart from historical data, live data about the COVID-19 number of infection cases, deaths, recovered individuals, and tests made is retrieved from the Worldometer web portal [37]. Although the information might not be taken as official as the one provided by John Hopkins University (which is based on WHO data), this source is updated many times per day providing the latest up-to-date data about COVID-19 statistics at all times. Thus its usefulness in providing the system with an almost real-time perspective on the pandemics worldwide.

3.3 Live news about coronavirus

The live news is retrieved from the global news engine Event Registry [16], which is a media-intelligence platform that collects news media from around the world in many languages. The service analyzes news from more than 30 thousand news, blog posts, and press releases daily, over around 75 thousand news sources in more than 60 languages. The multilingual capabilities of the system allow for the identification of topics across languages, ensuring a global coverage with local granularity [18]. With this data we are able to access what is the awareness of the media on aspects and key figures in the pandemics.

3.4 Google COVID-19 community mobility data

Google's Community Mobility [19] data compares mobility patterns that date from before the COVID-19 crisis, describing the situation on a weekly basis. Mobility patterns are measured as changes in the frequency of visits to six location types: retail and recreation; grocery and pharmacy; parks; transit stations; workplaces; and residential areas. The data is provided on a country level as well as on a province level. These reports are available for a limited amount of time, limited to their usefulness in supporting the work of public health officials. This data provides us with a wide perspective on the exchanges with potential of contamination based on mobility.

3.5 MEDLINE: medical research open dataset

In 2020 the MEDLINE dataset [23] contains more than 30 million citations and abstracts of the biomedical literature dating back to 1966. Over the past ten years, an average of a million articles were added each year. Around 5% of MEDLINE is on published research on infections, with cancer research being the most prevalent occuping 12% of this body of knowledge. Most scientific articles in this dataset are hand-annotated by health experts using 16 major categories and a maximum of 13 levels of deepness. The labeled articles are hand-annotated by humans based on their main and complementary topics, and on the chemical substances that they relate to. It is widely used by the biomedical research community through the well-accepted search engine PubMed [29]. The richness of this data allows us for insight on aspects of the disease as well as approaches and best practices from the published research.

4 Coronavirus watch dashboard

The main layout of the dashboard displayed in figure 1 consists of two sides. On the left side, the dashboard is split into the summarized information on the incidence across countries, where a simple table of statistics is provided about countries along with the total numbers of cases, death cases, and recovered cases. On the right side, there is a navigation panel with tabs, each representing a functionality. Each of these tabs is a view on the pandemics, based on its own specific data sources, and answers some questions and provides insights about a certain type of data focusing a specific aspect.

4.1 Coronavirus data table

The data table functionality is a straightforward data visualisation that shows the basic statistics about the new coronavirus. It's sourced from Worldometer as it's the most frequently updated source for coronavirus. This data table extends the one that is consistently shown in most views, a summarized version, to the table on the left containing the full information details. These include the new infection cases and the new death cases, in the day of the visit to the portal, as well as the total number of recovered cases, the active cases and the critical cases. The cumulative counts include the total number of infected, dead and tests. The latter are provided in absolute and per one million capita, taking into consideration the different population per country.

4.2 Coronavirus live news

The news monitoring functionality offers a live news feed about coronavirus-related topics from sources around the world. The feed is provided by the multilingual news engine Event Registry over an API. The ingested news article sample is generated by querying the system for articles that are annotated with concepts and keywords related to the coronavirus. The user can check for a country's specific news (based on the news source in that country) by clicking on the country name on the left table, as seen in figure 1.

This feature is a differentiator to most COVID-19 observatories, allowing the visitor to access real-time information about the pandemics, either at a global scale, in the European context or at country level. It also allows for a perspective on the public awareness on aspects of the pandemics.

4.3 Exploratory analytics

The following set of data visualization modules aim at displaying the statistical data about COVID-19 cases and deaths. While they all provide countries comparison, each one focus on a different perspective. Some are more complex and focused on the big picture (5D evolution), and some others are simple and focus on a unique aspect (Progression and Trajectory). Moreover, all of them have configuration options to tweak the visualization, like the ability to change the scale of the axes to focus on the top countries, or to focus on the long tale. Some offer a slider to manually move through the days for further inspection. Furthermore, the default view compares all the countries or the top N countries, depending on the visualization selected. However, it is possible to track a single country or a set of countries and compare them together for a more specific view. This is done when selecting the main country by clicking on it on the left table and proceeding to select more countries by pressing the ctrl key while clicking on the additional country.

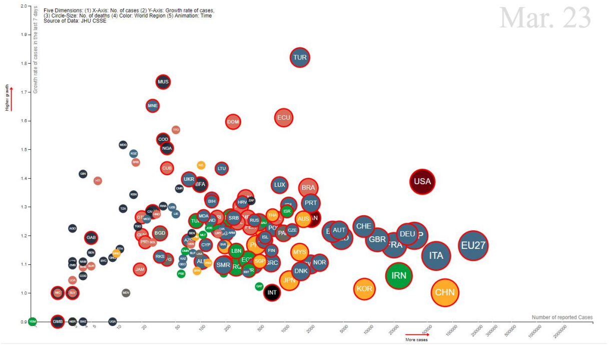


Figure 2: A snapshot of the 5D Visualization on March 23rd. Countries that were at the peak in terms of growth are shown high up.

4.3.1 5D evolution

5D Evolution is a visualization that displays the incidence of the virus in the population through time. It is named 5D since it encompasses five dimensions: x-axis, y-axis, bubble size, bubble color, and time, as seen in figure 2. By default, it illustrates the evolution of the pandemic in countries based on N cases (x-axis). The growth factor of N Cases (y-axis), N Deaths (bubble size), and country region (bubble color) through time. In addition, a red ring around the country bubble is drawn whenever the first death appears. The growth rate represents how likely the numbers are increasing with respect to the day before. A growth rate of 2 means that the numbers are likely to double in the next day. The growth rate is calculated using the exponential regression model. At each day the growth rate is based on the N cases from the previous seven days. The goal of this visualization is to show how countries relate to each other, which are "exploding in numbers" and which ones managed to "flatten the curve" (since flattening the curve means less growth rate). It is intended to be one visualization that gives the user a big picture of the situation.

4.3.2 Progression

The progression visualization displays the line graph comparing Date vs N cases/deaths. It helps to provide a simplistic view of the situation and compare countries based on the raw numbers only. The user can display the cumulative numbers where each day represents the current counts, or daily where at each date we count the cases/deaths on that day only.

4.3.3 Trajectory

While the progress visualization displays the normal date vs N cases/deaths, this visualization seeks to compare how the trajectory of the countries differ starting from the point where they detect cases. This visualization helps to compare countries' situations if they all start having cases on the same date. The starting point has been set to the day the country reaches 100 cases, so we would compare countries when they started gaining momentum.

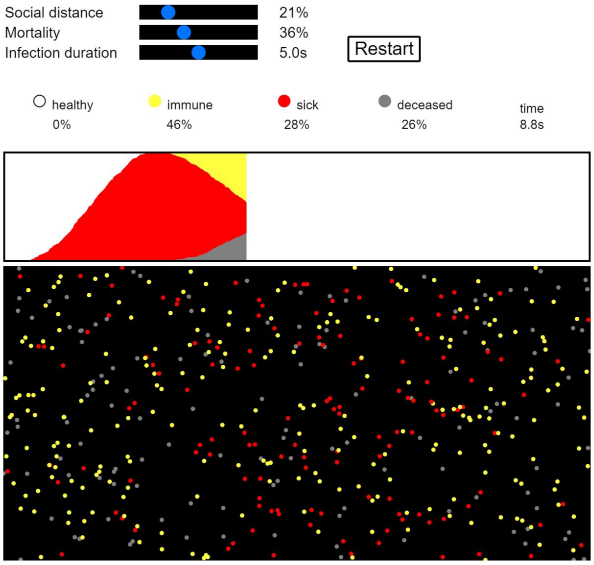
4.4 Time gap

The time gap functionality tries to estimate how the countries are aligned and how many days each country is behind the other, whether that is in the number of cases or deaths. This assumes that the trajectory of the country will continue as is without taking much more strict/loose measurements, which is a rough assumption. It helps to estimate how bad or good is the situation in terms of the number of days. To see the comparison, a country has to be selected from the table on the left. However, not all countries are comparable, having very different trajectories or growth rates.

The growth of each country is represented as an exponential function, the base is calculated using linear regression on the log of the historical values (that is, exponential regression). Based on that, the duplication N days, or the N days representing the number of cases/deaths will double is determined. Two countries are comparable if they have a reasonable difference in the base or doubling factor. If they are comparable, we see where the country with the smaller value fits in the historical values of the country with the larger numbers. We use linear interpolation if the number is not exact, hence the decimal values.

4.5 Mobility

The mobility visualization is based on google community mobility data that describe how communities in each country are moving based on 6 parameters: retail and recreation, grocery and pharmacy, parks, transit stations, workplaces, and residential areas. The data is then reduced to 2-dimensional data while keeping the Euclidean proximity nearly the same. The visualization can indicate that the closer the countries are on the visualization, the similar the mobility patterns they have. The visualization uses the T-SNE algorithm for dimensionality reduction [35], which reduces high dimensional data to low dimensional one while keeping the distance proximity between them proportionally the same as possible. The algorithm works in the form of iterations, at each iteration, the bubbles representing the country are drawn. We used those iterations to provide animation to the visualization.



Social distance 21%
Mortality 36%
Infection duration 5.0s

Figure 3: A snapshot of the Social Distancing Simulator. The canvas show a representation of the population. with red dots representing sick people, yellow dots representing immunized people, and grey dots represent deceased people.

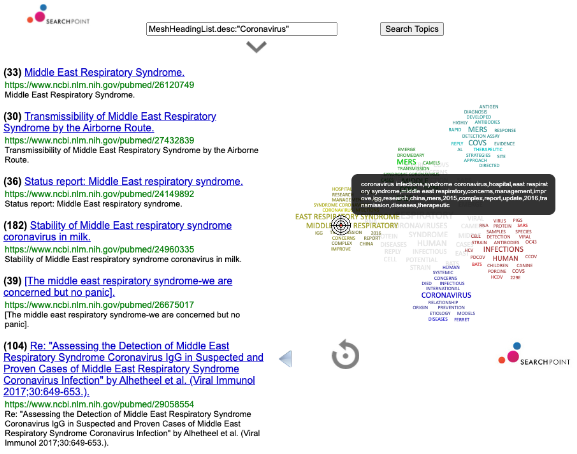
4.6 Social distancing simulator

The Social Distancing simulator is displayed in figure 3. Each circle represents a person who can be either healthy (white), immune (yellow), infected (red), or deceased (gray). A healthy person is infected when they collide with an infected person. After a period of infection, a person either dies or becomes permanently immune. Thus the simulation follows the Susceptible-Infectious-Recovered-Deceased (SIRD) compartmental epidemiological model.

The simulator is controlled by three parameters. Firstly, the Social distancing that controls to what extent the population enforces social distancing. At 0% there is no social distancing and persons move with maximum speed so that there is a great deal of contact between them. At 100% everyone remains still and there is no contact at all. Secondly, mortality is the probability that a sick person dies. If you set mortality to 0% nobody dies, while the mortality of 100% means that anybody who catches the infection will die. Finally, the infection duration determines how long a person is infected. A longer time gives an infected person more opportunities to spread the infection. Since the simulation runs at high speed, time is measured in seconds.

4.7 Biomedical research explorer

To better understand the disease, the published biomedical science is the source that provides accurate and validated information. Taking into consideration a large amount of published science and the obstacles to access scientific information, we made available a MEDLINE explorer where the user can query the system and interact with a



MeshHeadingList.desc:"Coronavirus"

(33) Middle East Respiratory Syndrome.
https://www.ncbi.nlm.nih.gov/pubmed/26120749
Middle East Respiratory Syndrome.

(30) Transmissibility of Middle East Respiratory Syndrome by the Airborne Route.
https://www.ncbi.nlm.nih.gov/pubmed/27432839
Transmissibility of Middle East Respiratory Syndrome by the Airborne Route.

(36) Status report: Middle East respiratory syndrome.
https://www.ncbi.nlm.nih.gov/pubmed/24149892
Status report: Middle East respiratory syndrome.

(182) Stability of Middle East respiratory syndrome coronavirus in milk.
https://www.ncbi.nlm.nih.gov/pubmed/24960335
Stability of Middle East respiratory syndrome coronavirus in milk.

(39) [The middle east respiratory syndrome-we are concerned but no panic].
https://www.ncbi.nlm.nih.gov/pubmed/26675017
[The middle east respiratory syndrome-we are concerned but no panic].

(104) Re: "Assessing the Detection of Middle East Respiratory Syndrome Coronavirus IgG in Suspected and Proven Cases of Middle East Respiratory Syndrome Coronavirus Infection" by Alhetheel et al. (Viral Immunol 2017;30:649-653.).
https://www.ncbi.nlm.nih.gov/pubmed/29058554
Re: "Assessing the Detection of Middle East Respiratory Syndrome Coronavirus IgG in Suspected and Proven Cases of Middle East Respiratory Syndrome Coronavirus Infection" by Alhetheel et al. (Viral Immunol 2017;30:649-653.).

coronavirus infections,syndrome coronavirus,hospital,east respiratory syndrome,middle east respiratory,concerns,management,improve,igg,research,china,mers,2015,complex,report,update,2016,transmission,diseases,therapeutic

Figure 4: The interactive MEDLINE data exploration showing a scientific article originally positioned as 182nd, now in the 4th place.

pointer to specify the search results (e.g., obtaining results on biomarkers when searching for articles hand-annotated with the MeSH class "Coronavirus").

To allow for the exploration of any health-related texts (such as scientific reports or news) we developed an automated classifier [8] that assigns to the input text the MeSH classes it relates to. The annotated text is then stored in Elasticsearch [27], from where it can be accessed through Lucene language queries, visualized over easy-to-build dashboards, and connected through an API to the earlier described explorer (see figure 4 and read [12], [30] and [26] for more detail).

The integration of the MeSH classifier with the worldwide news explorer Event Registry allows us to use MeSH classes in the queries over worldwide news promoting an integrated health news monitoring [13] and trying to avoid bias in this context [11]. An obvious limitation is a fact that the annotation is only available for news written in the English language, being the unique language in MEDLINE.

5 Conclusion and future work

In this paper, we presented the coronavirus watch dashboard as a use-case of observing pandemic. However, this methodology can be applied to other kinds of diseases given the availability of similar data. Having this global system in place we are collecting data on the usual numbers accounted on the pandemics by other platforms (as in e.g., Worldometer [37] or ECDC [6]) but within an integrated approach (as e.g. in the OxCOVID19 project [34]) aiming for data interoperability enhancing the business intelligence generated by the opservatory For further development, we plan to implement a local dashboard for other countries as well, which would provide local data in the local language. This would provide the local coverage (as in, e.g., Slednik [32]) that can bring the global problems addressed to a specific and more usefu context (as was initiated in the context of the MIDAS project [12]). In addition, given the existence of more than seven months of historical data, we would like to build some predictive models to predict the number of cases/deaths in the next days.

Moreover, we are using the StreamStory technology [33] [14] in order to: (i) compare the evolution of the disease between countries by comparing their time-series of incidence; (ii) investigate the correlation between the incidence of the disease with weather conditions and other impact factors; and (iii) analyze the dynamics of the evolution of the disease based on incidence, morbidity, and recovery. This technology allows for the analysis of dynamical Markov processes, analyzing simultaneous time-series through transitions between states, offering several customization options and data visualization modules.

Furthermore, following the work done in the context of the Influenza epidemic in [9], we are using Topological Data Analysis methods in [10] to understand the behavior of COVID-19 in the European context [10] through the behaviour of the data it is reflected in. This is done analysing simultaneously a variety of time-series representing the several aspects of the pandemics, in a high dimension space. In it, we examine the structure of data through its topological structure, which allows for comparison of the evolution of the epidemics within countries through the encoded topology of their incidence time series.

Acknowledgement

The first author has been supported by the Knowledge 4 All foundation and the H2020 Humane AI project under the European research and innovation programme under GA No. 761758), while the second author was funded by the European Union research fund 'Big Data Supporting Public Health Policies', under GA No. 727721. The third author acknowledges that this material is based upon work supported by the Air Force Office of Scientific Research under award number FA9550-17-1-0326.

References

[1] H. Allahverdipour. 2020. Global challenge of health communication: infodemia in the coronavirus disease (covid-19) pandemic. *J Educ Community Health*, 7, 2, 65.–67.

[2] ArcGIS. 2020. ArcGIS who covid-19 dashboard. https://covid19.who.int/. (2020).

[3] European Commission. [n. d.] The european covid-19 data portal. https://www.covid19dataportal.org/. Accessed in: April 2021. ().

[4] CoronaTracker. 2020. CoronaTracker. https://www.coronatracker.com/analytics/. (2020).

[5] CSSE. 2020. Covid-19 data repository by the center for systems science and engineering (csse) at johns hopkins university. `https://github.com/CSSEGISandData/COVID-19`. (2020).

[6] ECDC. [n. d.] Ecdc covid-19 datasets. `https://www.ecdc.europa.eu/en/covid-19/data`. Accessed in: April 2021. ().

[7] B. Xu et al. 2020. Epidemiological data from the covid-19 outbreak, real-time case information. *Scientific data*, 7, 1, 1–6. DOI: `10.1038/s41597-020-0448-0`.

[8] J. Pita Costa et al. 2021. A new classifier designed to annotate health-related news with mesh headings. *Artificial Intelligence in Medicine*, 114, 102053. DOI: `10.1016/j.artmed.2021.102053`.

[9] J. Pita Costa et al. 2019. A topological data analysis approach to the epidemiology of influenza. In *Proceedings of the Slovenian KDD conference*.

[10] J. Pita Costa et al. 2020. A topological data analysis perspective on the covid-19 pandemics. In *In preparation*.

[11] J. Pita Costa et al. 2019. Health news bias and its impact in public health. In *Proceedings of the Slovenian KDD conference*.

[12] J. Pita Costa et al. 2020. Meaningful big data integration for a global covid-19 strategy. *Computer Intelligence Magazine*, 15, 4, 51–61. DOI: `10.1109/MCI.2020.3019898`.

[13] J. Pita Costa et al. 2017. Text mining open datasets to support public health. In *WITS 2017 Conference Proceedings*.

[14] L Stopar et al. 2018. Streamstory: exploring multivariate time series on multiple scales. *IEEE transactions on visualization and computer graphics*, 25, 4, 1788–1802. DOI: `10.1109/TVCG.2018.2825424`.

[15] T. Kuchler et al. 2020. The geographic spread of covid-19 correlates with the structure of social networks as measured by facebook. *National Bureau of Economic Research*, w26990. DOI: `10.1016/j.jue.2020.103314`.

[16] EventRegistry. 2020. Event Registry. `https://eventregistry.org`. (2020).

[17] Center for Disease Control. 2020. Center for Disease Control and Prevention. `https://www.cdc.gov/coronavirus/2019-ncov/index.html`. (2020).

[18] J. Brank G. Leban B. Fortuna and M. Grobelnik. 2014. Event registry: learning about world events from news. In *In Proceedings of the 23rd International Conference on World Wide Web*, 107–110. DOI: `10.1145/2567948.2577024`.

[19] Google. 2020. Google COVID-19 Community Mobility Report. `https://www.google.com/covid19/mobility/`. (2020).

[20] IEEE. [n. d.] Ieee xplore covid-19 resources. `https://ieeexplore.ieee.org/Xplore/home.jsp`. Accessed in: April 2021. ().

[21] IRCAI. 2020. IRCAI coronavirus watch portal. `http://coronaviruswatch.ircai.org/`. (2020).

[22] Kaggle. 2020. Kaggle covid-19 open research dataset challenge. `https://www.kaggle.com/allen-institute-for-ai/CORD-19-research-challenge`. (2020).

[23] MEDLINE. 2020. MEDLINE description of the database. `https://www.nlm.nih.gov/bsd/medline.html`. (2020).

[24] medRxiv. [n. d.] Covid-19 sars-cov-2 preprints from medrxiv and biorxiv. `https://connect.medrxiv.org/relate/content/181`. Accessed in: April 2021. ().

[25] MeSHNow. 2020. MeSHNow. `https://www.ncbi.nlm.nih.gov/CBBresearch/Lu/Demo/MeSHNow/`. (2020).

[26] MIDAS. 2020. MIDAS COVID-19 portal. `http://www.midasproject.eu/covid-19/`. (2020).

[27] Elastic NV. 2020. Elasticsearch portal. `https://www.elastic.co/`. (2020).

[28] World Health Organisation. 2020. WHO Coronavirus portal. `https://www.who.int/emergencies/diseases/novel-coronavirus-2019`. (2020).

[29] PubMed. 2020. PubMed biomedical search engine. `https://pubmed.ncbi.nlm.nih.gov/`. (2020).

[30] Quintelligence. 2020. Quintelligence COVID-19 portal. `http://midas.quintelligence.com/`. (2020).

[31] Ravenpack. 2020. Ravenpack coronavirus news monitor. `https://coronavirus.ravenpack.com/`. (2020).

[32] Slednik. [n. d.] Covid-19 sledilnik slovenija. `https://www.sledilnik.org`. Accessed in: April 2021. ().

[33] Luka Stopar. 2020. StreamStory. `http://streamstory.ijs.si/`. (2020).

[34] Oxford University. [n. d.] Oxford covid-19 (oxcovid19) project. `https://covid19.eng.ox.ac.uk/`. Accessed in: April 2021. ().

[35] Laurens van der Maaten and Geoffrey Hinton. 2008. Viualizing data using t-sne. *Journal of Machine Learning Research*, 9, (November 2008), 2579–2605.

[36] WHO. 2020. World Health Organization who director-general's opening remarks at the media briefing on covid-19 - 11 march 2020. https://www.who.int/dg/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020. (2020).

[37] WorldoMeters. 2020. WorldoMeters. https://www.worldometers.info/coronavirus/. (2020).

[38] Zenodo. [n. d.] Zenodo coronavirus disease research community - covid-19. https://zenodo.org/communities/covid-19/. Accessed in: April 2021. ().