

CROSS-LANGUAGE SEMANTIC PERSPECTIVE ENABLING CROSS-CULTURAL COMPARISONS OF PEOPLE’S REACTIONS ON COVID-19 OUTBREAK ON SOCIAL MEDIA

KNOWLEDGE EXTRACTION FROM BIG DATA

IN DIGITAL HEALTH

ROSSANA COTRONEO – ENEA Direzione Trasferimento Tecnologico rossana.cotroneo@enea.it | DANIELA ALDERUCCIO – ENEA ICT-HPC Lab daniela.alderuccio@enea.it

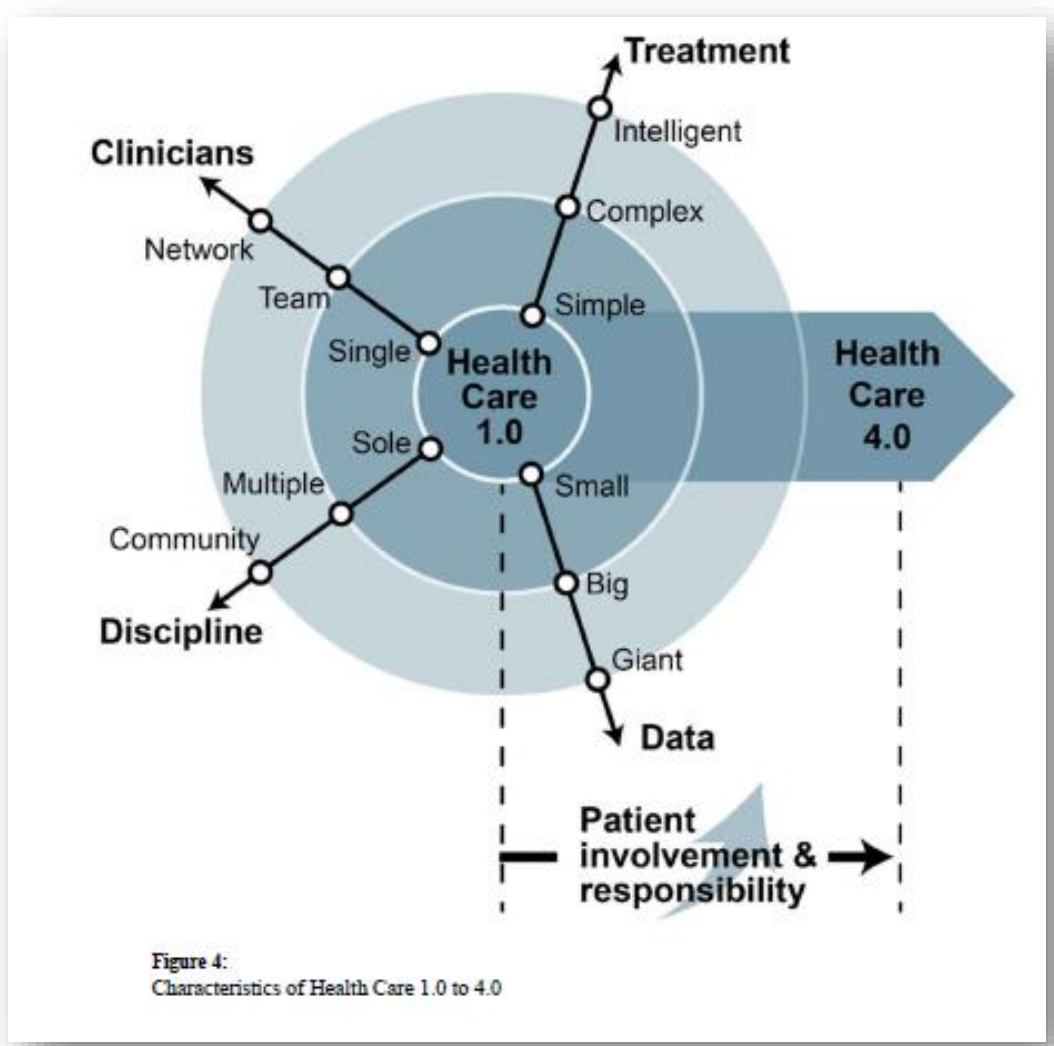
Abstract. The aim of this study is to extract re-usable knowledge from open sources, supporting prediction and prescriptive models in healthcare sectors, for a healthier Society (GOAL 3) Description We process e_Health multilingual data collections and extract specific Knowledge and keywords from tweets during Covid-19 pandemic, then design a pipeline for processing and analyzing data. Results We adopt a cross-linguistic semantic perspective, refining crawling/analysis strategies and enabling cross-cultural comparisons of people's reactions on coronavirus outbreak on social media. We also formulate hypothesis on the dynamics of information/knowledge spreading: (i) from institutional hub to users/patients; (ii) and among patients.

Expected growth of Digital Healthcare Data

The Digital Healthcare data in 2012 was estimated to be equal to 500 petabytes and in 2020 it was expected to be 25,000 petabytes. (Sun & Reddy, 2013).

In 2022 Worldwide digital healthcare data is estimate to currently equal between 25 exa-bytes (25 × 10¹⁸ bytes) and 35 zeta-bytes (35 × 10²¹ bytes) with an annual increase of between 1.2 and 2.4 exabytes per year

Source: Šin P, Hokyňková A, Marie N, Andrea P, Křt R, Podroužek J. Machine Learning-Based Pressure Ulcer Prediction in Modular Critical Care Data. Diagnostics. 12(4):850. https://doi.org/10.3390/diagnostics12040850 – IEEE Big Data. Available online: https://bigdata.ieee.org/ (accessed on Feb 2022).



Such a huge amount of patient data is generated by a variety of lab systems and health information systems (e.g., EHRs – Electronic Health Records). Factors driving the global digital health market and augmenting its growth are:

- the growing demand for mobile health apps,
- the increasing demand for remote patient monitoring services,
- the rising adoption of smartphones and tablets.

New Data Ecosystem integrates traditional and modern data sources (i.e. social media platforms). During the SARS-COV2 Pandemy Twitter witnesses its impact on citizens lives, offering insights into user health status, symptoms, concerns and opinions. Although Twitter represents an alternative platform for infodemiological investigations, tweet frequency on COVID-19–related symptoms was similar in sensitivity to Google Search trends (Panuganti et al, 2020); furthermore, tracking public health information from online search engines (“infoveillance”) complements traditional public health surveillance systems (Ciaffi et al., 2020)

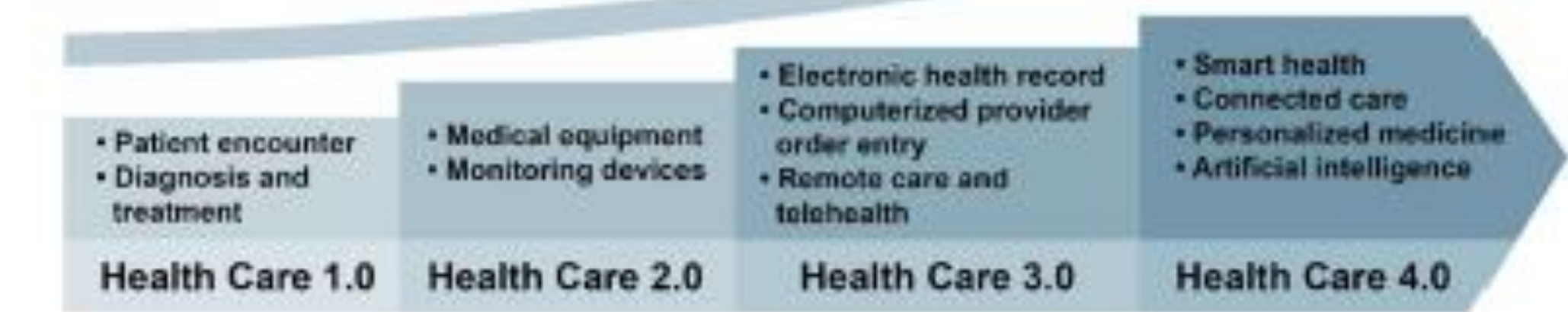
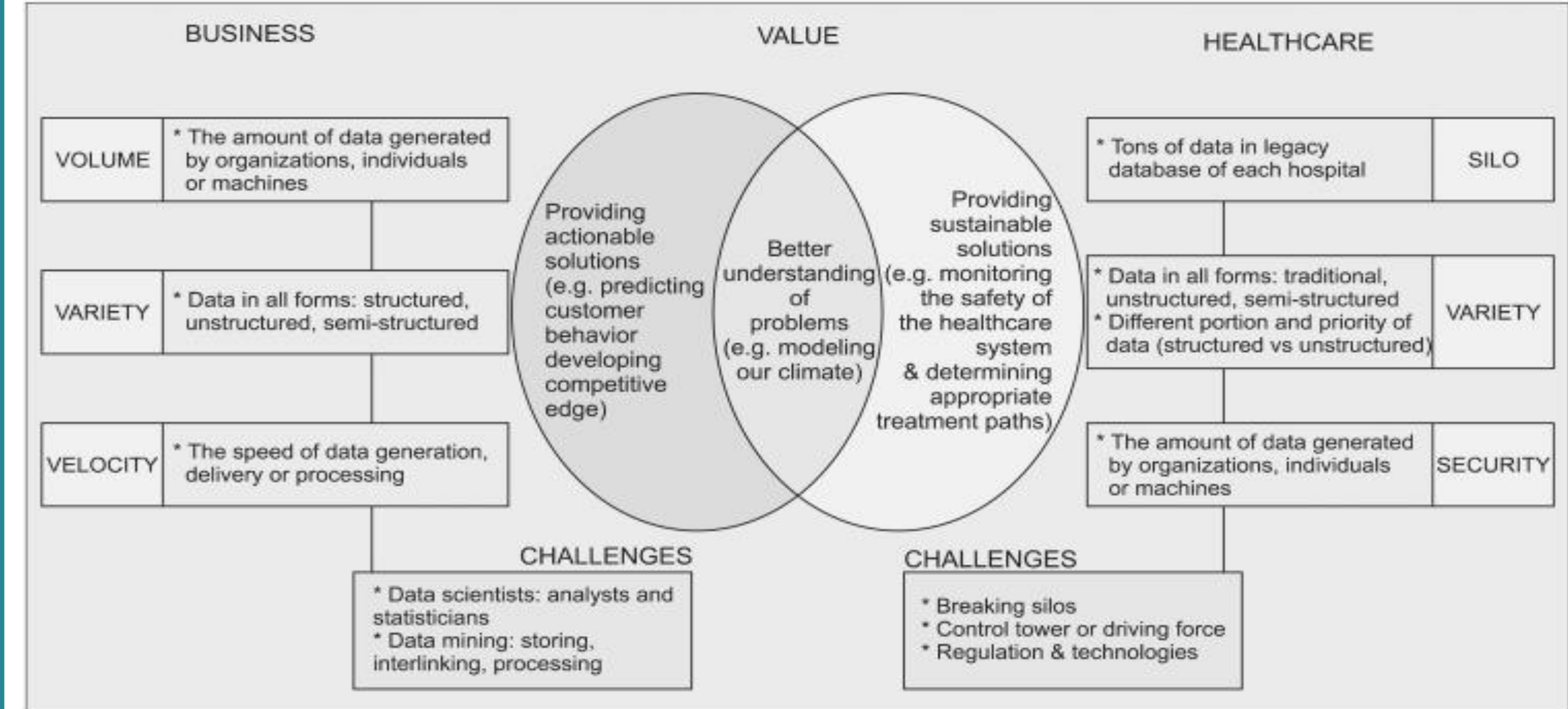


Figure 3. Historical Evolution of Health Care 1.0 to Health Care 4.0
*Source: Li, J. - Carayon, P. Health Care 4.0: A vision for Smart and Connected Health Care. IISE Trans Healthc Syst Eng 2021; 11(3); 171-180. doi 10.1080/24725579.2021.1884627



<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3717441/>



In this study we process e-Health data from tweets collections and extract specific Knowledge and keywords in 2020 during the initial phase of the pandemy in Italy. On March 20th, 2020 we searched for the keyword to be used to select tweets in Twitter Trends in Italy in the period from 16th to 18 March and extracted keywords (see Table) containing virus denomination and sometimes geolocalization, only geolocalization, or keyword linked to: the pandemic emergency; governmental measure; social activities /actions supporting people; negative emotions; disease, protection devices and recommended behaviour; governmental rules: political figures (pro/cons) and other keywords without any semantic correlations, that apparently did not have nothing in commons with the pandemic (i.e. #leCose - #iorestocasa). We will show a pipeline of sequential steps to be performed, in order to process and analyse dynamic continuous data. A generic data processing pipeline has following steps: 1st step: Exploring the Semantic field (NLP) for Keyword Identification – 2nd Step: Data Crawling of topic and Raw Data Storage - 3rd Step: Data Gathering & NLP Exploration / Analysing/Synthesising/Visualizing/Results' storage. In this on-going research, we :

(i) define a pipeline to process Healthcare data, from raw data stream ingestion → data stream pre-processing → data stream processing → stream output & visualization (data stream storage: stores, indexes, manages data and Knowledge); due to the domain-specific nature of big data in the healthcare sector, it is important to select appropriate methods in data gathering, pre-processing, data analysis, interpretation and visualization,

(ii) and formulate hypothesis on the dynamics of information/news spreading from hub to users(patient), from the centre to the periphery.

Twitter Trending Topics Evolution from 16th-18th March, 2020

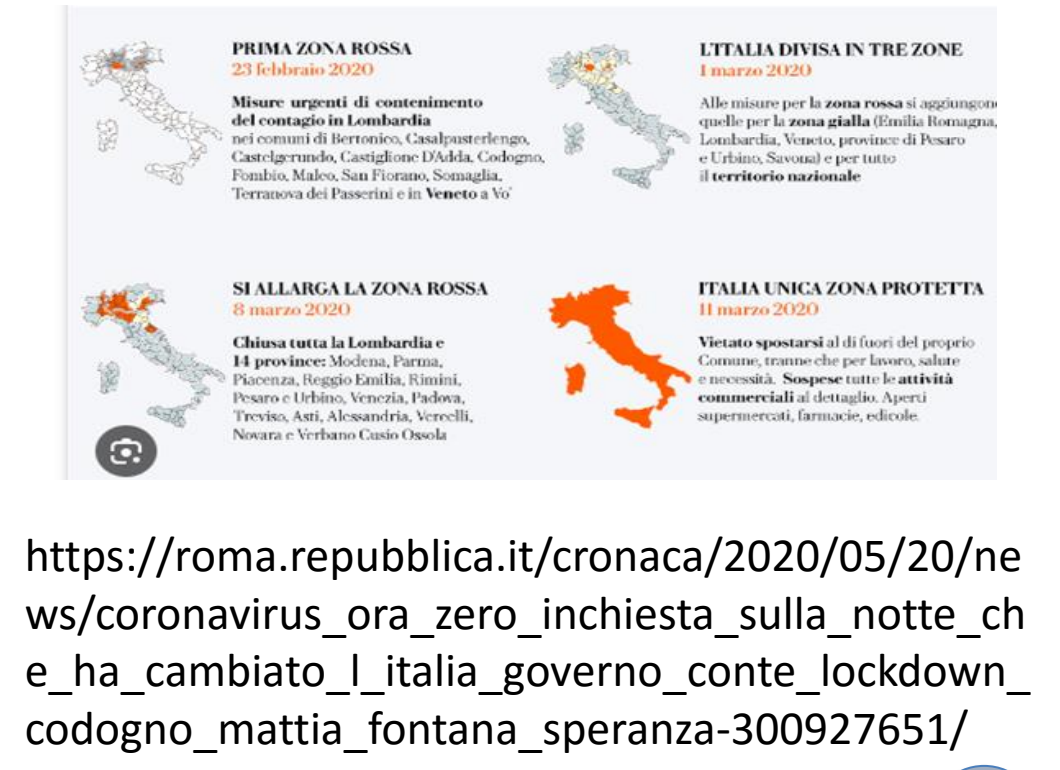
| Topics 16 th -18 th March 2020 | 16th March 2020 | 18th March 2020 |
|-------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------|
| <u>virus denomination and geolocalization</u> | #coronavirusupdates - #CoronaVirusUpdates #COVID19italia | #COVID19italia #coronavirusitalia - #CoronaVirusitaly - #Bergamo |
| <u>governmental measure</u> | #Cura Italia | #decretocuraitalia - #CuraItalia |
| <u>social activities in support of people and emotion</u> | #orgoglioitaliano - #DomaniUsciraiSole - #flashmob - #Orgoglioitaliano - #celafaremo - #flashmobsonoro - #flashmobitalia - #balconi - #flashmobsonoro - #PaesaggiDaCartolina - #stayhomechallenge | #CoronaVirusChallenge - #flashmob - #flashmobitalia - #balconi - #flashmobsonoro - #PaesaggiDaCartolina - #stayhomechallenge |
| <u>disease, protection devices and recommended behavior</u> | #scriverquelchesento - #DomaniUsciraiSole #coronapocalypse | #quarantinelifelife #18marzo - #quarantinelifelife - #coronapocalypse - #Mascherine - #mascherine - #decretocuraitalia - #CuraItalia #unitaditalia |
| <u>politicians(pro/cons) and national support/solidarity, unity of people</u> | e Luca - #boris johnson - #GovernoConte - #governoconte - #loStoconConte - #contedimettiti - #CoronaVirusitaly | |

Source: Elaboration from Twitter Trending Topics (16th-18th March,2020)

Twitter Trending Topics

Twitter trends on 17th March
#emergenzacovid19 - #curaitalia - #EmergenzaCovid19)

Twitter trends on 18th March
#quarantinelifelife - #QuarantineLife - #decretocuraitalia - #COVID19italia - #CuraItalia
show also covid-related hashtag mixed with other topic:
#18Marzo worldwide twitter trend - #leCose - #PaesaggiDaCartolina - Fabio Fazio - #IncantoEpoesia - #Ascierto - buongiorno ferula - Vanessa Hudgens - tom riddle - hermione - enzo miccio - hogwarts - Dobby - silente - Gennaro - draco - wendy - hagrid - Sharpay - voldemort - enzo e vera - gilderoy allock - paura potter - harry e ron - Coachella - Silvio - Serpeverde - mirilla malcontenta - lucius malfoy - Dayane - Legnano - Ginny - forza blaise - weasley - #Mentana - #HarryPotter - #pechinoexpress - #QuarantineLife - #CuraItalia - #dimartedi - #unitaditalia - #fuoridalcoro - #primoappuntamento - #euro2021 - #stayhomechallenge - #terremoto - #Matuidi - #Limonov - #armymissesbts - #sanpatrizio



Testo Twitter 7:30 PM · 18 mar 2020 :
“Ti auguro di riprenderti presto e grazie per la tua testimonianza che spero serva a quei ragazzi che vogliono sfidare la sorte!!! E non solo ai ragazzi!!! #leCose non cambiano, se non cambiamo noi. #CoronaVirusChallenge #coronavirus”

Results show that the adoption of a cross-language semantic perspective to multilingual data is a propaedeutic step able to refine both crawling and analysis strategies. Furthermore, it enables the detection of national and cross-cultural trends in facing a global crisis as emerged in tweets in the initial phase of the coronavirus pandemic. Big Data Analytics impacts on governments, organizations and health facilities, to draw strategic planning, providing insights on the disease conditions in a region and the behaviour of people in the area. The cross-cultural comparison of people's reactions to extreme event provides insights into different collective cross-country reactions on coronavirus outbreak on social networks. Exploring cultural similarities and differences of behaviour against a social crisis (epidemy, earthquake, extreme climate events, etc.) will help in developing appropriate national policy and social acceptance, and supporting prediction and prescriptive models in healthcare sectors, for a healthier Society (GOAL 3)