Lest We Forget: A Dataset of Coronavirus-Related News Headlines in Swiss Media

Alireza Ghasemi  
ELCA Informatik AG  
Zürich, Switzerland  
alireza.ghasemi@elca.ch

Amina Chebira  
ELCA Informatique SA  
Lausanne, Switzerland  
amina.chebira@elca.ch

Abstract—We release our COVID-19 news dataset, containing more than 10,000 links to news articles related to the Coronavirus pandemic published in the Swiss media since early January 2020. This collection can prove beneficial in mining and analysis of the reaction of the Swiss media to the COVID-19 pandemic and extracting insightful information for further research.

We hope this dataset helps researchers and the public deliver results that will help analyse the pandemic and potentially lead to a better understanding of the events.

Index Terms—COVID-19, Natural language processing, Data mining, Text analytics

I. INTRODUCTION

The COVID-19 pandemic started in Switzerland on February 25th 2020, when the first infection was officially reported in the Italian-speaking canton of Ticino [1], [2]. Soon the pandemic spread around the country on all cantons, and Switzerland became one of the most infected countries on a per-capita basis [3], [4].

The Swiss government started putting in place various measures to control and suppress the pandemic. Gatherings were limited and later totally banned, following by closure of all except essential business, and finally closing land borders with neighbouring countries [5].

These measures helped control the spread of the virus and significantly decreased the number of active and daily new cases in Switzerland. With the success confirmed, government started gradually lifting the established restrictions from late April [6]. Finally, the June 15th re-opening land borders with the neighbouring countries marked the “end” of the pandemic in Switzerland, at least for the first wave [7].

Since the first recorded case of the COVID-19 virus in Switzerland and far before as it was gaining attention around the world, The Swiss media started covering the topic from various aspects, including the everyday news about the state of the country, the immediate effects, and longer-term consequences of the pandemic. Given the multi-lingual and multi-cultural nature of Switzerland, interesting analyses can be accomplished to see how the media coverage of the pandemic has been managed and what topics in respect to the pandemic have been important to the Swiss media, and hopefully, by proxy to the Swiss public.

In order to help the research community and the public be able to analyse and seek answers to the above questions, we at ELCA decided to release our COVID-19 news dataset, containing more than 10,000 links to news articles related to the Coronavirus pandemic published in the Swiss media since early January 2020.

We hope this dataset helps researchers make insightful analyses on the reaction of the Swiss public to the pandemic and deliver results that help shape a better response in the prospective future cases.

II. THE DATA

We tried to cover the most popular Swiss newspapers and news websites. Therefore, we chose a total of 10 news sources in German, five in French, three in Italian, and also two English-speaking Swiss news websites in order to make the data more accessible to to researchers outside Switzerland. Table I depicts the list of venues and some information about them.

| News Source            | Language | Number of Articles |
|------------------------|----------|--------------------|
| Blick                  | German   | 2080               |
| Neue Zürcher Zeitung  | German   | 1961               |
| Tages Anzeiger         | German   | 1109               |
| Aargauer Zeitung       | German   | 1099               |
| Basler Zeitung         | German   | 859                |
| SRF                    | German   | 718                |
| 20 Minuten             | German   | 553                |
| Berner Zeitung         | German   | 281                |
| 20 Minutes             | French   | 225                |
| Tribune de Genve       | French   | 233                |
| Corriere del Ticino    | Italian  | 199                |
| 24 heures              | French   | 189                |
| World Radio Switzerland| English  | 176                |
| Il portale del Ticino  | Italian  | 156                |
| RTS                    | French   | 188                |
| SWI swissinfo.ch       | English  | 112                |
| Le temps               | French   | 89                 |
| SWI swissinfo.ch       | French   | 71                 |
| SWI swissinfo.ch       | German   | 55                 |
| SWI swissinfo.ch       | Italian  | 52                 |

A. Selection of Relevant Articles

We backdated our data collection to late 2019, and started scanning front pages of the selected news sources in consecutive days, extracting headlines of the articles. Initially, articles...
with any of the following *keywords* in the title were deemed "Coronavirus-related":
- Coronavirus,
- Covid,
- Lockdown,
- Pandem* (To account for different spellings of the concept in different languages).

This inevitably leads to false negatives. In order to reduce such false negatives, we read at a later stage the synopsis of the article and searched for the keywords also in the body, yielding more positive results. The distribution of the languages in the dataset is depicted in Figure 1

The first article we could find in the Swiss media has been published on January 8th in the French-speaking news portal 20 Minutes, titled "Un nouveau coronavirus apparat en Chine" [8]. We have made a web application to simplify exploring and browsing the data, and reading the collected news articles. The web application is available at https://covidnewsdataset.herokuapp.com/

**Summary**

We explained in this article our Swiss COVID-19 dataset and how it has been collected. We publish the dataset hereby for public use, along with an online visualisation application to help explore and look at the news articles of Swiss media during the pandemic in Switzerland. We hope this dataset proves useful in analysis of the pandemic era and the public response to it in Switzerland.

**References**

[1] The Local, “Breaking: Switzerland confirms first case of coronavirus,” February 25, 2020 (accessed June 21, 2020). [Online]. Available: https://www.thelocal.ch/20200225/breaking

[2] Blick, “Erster bestätigter Fall in der schweiz,” February 25, 2020 (accessed June 21, 2020). [Online]. Available: https://www.blick.ch/news/15767285

[3] 24 Heures, “Vaud enregistre 7 décès, 32 personnes aux soins intensifs,” March 20, 2020 (accessed June 21, 2020). [Online]. Available: https://www.24heures.ch/suisse/coronavirus/vaud-enregistre-7-deces-32-personnes-soins-intensifs

[4] The Local, “Coronavirus in Switzerland: Number of cases rises above 260,” March 7, 2020 (accessed June 21, 2020). [Online]. Available: https://www.thelocal.ch/20200303/more-cases-of-coronavirus-are-confirmed-in-switzerland

[5] Swiss Confederation, “Bundesrat verschärft massnahmen gegen das coronavirus zum schutz der gesundheit und unterstützt betroffene branchen,” March 14, 2020 (accessed June 21, 2020). [Online]. Available: https://www.admin.ch/gov/de/start/dokumentation/mediennachrichten.msg-id-78437.html

[6] RTE, “Switzerland announces gradual easing of COVID-19 restrictions,” April 16, 2020 (accessed June 21, 2020). [Online]. Available: https://www.rte.ie/news/coronavirus/2020/0416/1131999-switzerland-coronavirus

[7] SRF, “Wiedereröffnung der Grenzen: Das müssen Reisende aus der Schweiz jetzt wissen,” June 14, 2020 (accessed June 21, 2020). [Online]. Available: https://www.srf.ch/news/panorama/wiedereroffnung-der-grenzen-das-mussen-reisende

[8] 20 Minutes, “Un nouveau coronavirus apparat en Chine,” January 8, 2020 (accessed June 21, 2020). [Online]. Available: https://www.20min.ch/fr/story/355008750949