Collecting Verified COVID-19 Question Answer Pairs

Adam Poliak, Max Fleming, Cash Costello, Kenton W Murray, Mahsa Yarmohammadi, Shivani Pandya, Darius Irani, Milind Agarwal, Udit Sharma, Shuo Sun, Nicola Ivanov, Lingxi Shang, Kaushik Srinivasan, Seolhwa Lee, Xu Han, Smisha Agarwal, João Sedoc

https://covid-19-infobot.org/

JOHNS HOPKINS UNIVERSITY

BARNARD COLLEGE COLUMBIA UNIVERSITY

NYU | STERN

FAQ Dataset

Over 2,100 COVID-19 related Frequently asked Question-Answer pairs scraped from over 40 trusted websites

Scraping Criteria for Data Sources

1) Informativeness and trustworthiness 2) Ease of scraping frequently asked question-answer pair 3)Number of questions and answers

Example FAQ

```
"sourceUrl": "https://www.fda.gov/emergency-preparedness-and-response
              /mcm-issues/coronavirus-disease-2019-covid-19-frequently
              -asked-questions",
"sourceName": "FDA",
"typeOfInfo": "QA",
"hasAnswer": true,
"targetLocation": "US",
"language": "en",
"questionText": "If I vape tobacco or nicotine am I at risk for complications
              from COVID-19?",
"answerText": "E-cigarette use can expose the lungs to toxic chemicals, but
              whether those exposures  increase the risk of COVID-19 or the
              severity of COVID-19 outcomes is not known. However, many
              e-cigarette users are current or former smokers, and cigarette
              smoking increases the risk of respiratory infections,
              including pneumonia.",
"questionUUID": "cf4ffbaa-8a3d-11ea-a5a4-0af24cbcf56f",
"answerUUID": "cf4ffbab-8a3d-11ea-a5a4-0af24cbcf56f",
"ID": "FDA|||8643362802117439021",
"answerContainsURLs": false,
"answerToks2URL": {},
"dateLastChanged": 1588269805.917085
```

Paper   Data

Data Sources & Statistics

Source Name	# of Questions-Answers
North Dakota Stake Government	305
Vermont Department of Health	151
NYTimes	118
CNN	106
Kansas Department of Health and Enviroment	92
FDA	76
Oregon Public Health Division	75
Johns Hopkins Bloomberg School of Public Health	57
FloridaGov	45
Texas Human Resources	40
National Foundation for Infectious Diseases	33
AVMA	32
WHOMyth	29
Cleveland Clinic	29
Public Health Agency of Canada	28
Ministero della Salute, IT	16
JHU Medicine	13
JHU HUB	7
Hawaii State Government	4
Delaware State Government	1
GOV Polska	154
Bundesministerium für Gesundheit (BMG)	201
FHM, Folkhälsomyndigheten	142
Ministero della Salute, IT	16
World Health Organization (WHO)	121
Bundesministerium für Wirtschaft und Energie	34
Berliner Senat	48
European Centre for Disease Prevention and Control	47
Bundesanstalt für Arbeitsschutz und Arbeitsmedi...	35
Bundesministerium für Arbeit und Soziales (BMAS)	32
Bundesagentur für Arbeit	11
Presse- und Informationsamt der Bundesregierung	16
Center for Disease Control and Prevention (CDC)	13
Robert Koch Institute (RKI)	4
total	2115

Manually Aligning Questions and Answers

24,000 online questions aligned by public health experts with answered questions from our dataset.

Alignment Instructions

Protocol

There is a lot of subjectivity that enters the picture when we are assigning relevancy through the use of a 0-100% scale. To help standardize our thought process and understanding on how to rank relevancy, we've tried to do it by parsing out systematically what each question is discussing.

Please note that these examples are only showing it one at a time. As you can see in the description of the task and the picture above, it will actually be one user question, and five relevant questions to review

Example 1:

User Question: How many people have it in my town

Relevant Question: How long do people have to isolate for?

Thought Process: The relevant question is focused on COVID-19, isolation, and duration. The new question is focused on COVID-19 and prevalence.

Relevancy Scale: 0% relevant

- They are both talking about COVID-19, but don't overlap significantly beyond that.