Ad-Hoc Monitoring of COVID-19 Global Research Trends for Well-Informed Policy Making

SOUVIKA SARKAR, Auburn University, USA
BIDDUT SARKER BIJOY and SYEDA JANNATUS SABA, Shahjalal University of Science and Technology, Bangladesh
DONGJI FENG and YASH MAHAJAN, Auburn University, USA
MOHAMMAD RUHUL AMIN, Fordham University, USA
SHEIKH RABIUL ISLAM, University of Hartford, USA
SHUBHRA KANTI KARMAKER ("SANTU"), Auburn University, USA

The COVID-19 pandemic has affected millions of people worldwide with severe health, economic, social, and political implications. Healthcare Policy Makers (HPMs) and medical experts are at the core of responding to this continuously evolving pandemic situation and are working hard to contain the spread and severity of this relatively unknown virus. Biomedical researchers are continually discovering new information about this virus and communicating the findings through scientific articles. As such, it is crucial for HPMs and funding agencies to monitor the COVID-19 research trend globally on a regular basis. However, given the influx of biomedical research articles, monitoring COVID-19 research trends has become more challenging than ever, especially when HPMs want on-demand guided search techniques with a set of topics of interest in mind. Unfortunately, existing topic trend modeling techniques are unable to serve this purpose as (1) traditional topic models are unsupervised, and (2) HPMs in different regions may have different topics of interest that they want to track.

To address this problem, we introduce a novel computational task in this article called *Ad-Hoc Topic Tracking*, which is essentially a combination of *zero-shot* topic categorization and the spatio-temporal analysis task. We then propose multiple *zero-shot* classification methods to solve this task by building on state-of-the-art language understanding techniques. Next, we picked the best-performing method based on its accuracy on a separate validation dataset and then applied it to a corpus of recent biomedical research articles to track COVID-19 research endeavors across the globe using a spatio-temporal analysis. A demo website has also been developed for HPMs to create custom spatio-temporal visualizations of COVID-19 research trends. The research outcomes demonstrate that the proposed *zero-shot* classification methods can potentially facilitate further research on this important subject matter. At the same time, the spatio-temporal visualization tool will greatly assist HPMs and funding agencies in making well-informed policy decisions for advancing scientific research efforts.

Authors' addresses: S. Sarkar, D. Feng, Y. Mahajan, and S. K. Karmaker ("Santu"), 3106 Shelby Center, College of Engineering, Auburn University, Auburn, AL 36849, United States of America; emails: szs0239@auburn.edu, dzf0023@auburn.edu, yzm0034@auburn.edu, sks0086@auburn.edu; B. S. Bijoy and S. J. Saba, Shahjalal University of Science and Technology Kumargaon, Sylhet, Sylhet, 3114, Bangladesh; emails: biddut12@student.sust.edu, syeda06@student.sust.edu; M. R. Amin, Fordham University, Department of Computer and Information Sciences, 113 W 60th St, New York, NY 10023, United States, email: mamin17@fordham.edu; S. R. Islam, University of Hartford, Department of Computing Sciences, 200 Bloomfield Ave, West Hartford, CT 06117, USA; email: shislam@hartford.edu.

2157-6904/2023/02-ART26 $15.00
https://doi.org/10.1145/3576901

CCS Concepts: • **Information systems** → **Users and interactive retrieval**; • **Applied computing** → **Document searching**; • **Computing methodologies** → **Information extraction;**

Additional Key Words and Phrases: Topic models, zero-shot learning, COVID-19, policy making, spatio-temporal analysis

ACM Reference format:
Souvika Sarkar, Biddut Sarker Bijoy, Syeda Jannatus Saba, Dongji Feng, Yash Mahajan, Mohammad Ruhul Amin, Sheikh Rabiul Islam, and Shubhra Kanti Karmaker ("Santu"). 2023. Ad-Hoc Monitoring of COVID-19 Global Research Trends for Well-Informed Policy Making. *ACM Trans. Intell. Syst. Technol.* 14, 2, Article 26 (February 2023), 28 pages.
https://doi.org/10.1145/3576901

1 INTRODUCTION

The COVID-19 pandemic, the spread of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), has led to a dramatic loss of human lives worldwide and presented an unprecedented challenge to prevent further loss of life as well as maintain livelihoods. The economic and social disruption caused by the pandemic has been devastating, rapidly affecting our day-to-day life and businesses as well as disrupting world trade, movements, and, in turn, the global economy. Healthcare Policy Makers (HPMs) are increasingly pressed to articulate their rationales and strategies for containing the COVID-19 pandemic. As the trade-off between further disease spread and socioeconomic costs is debated, it is essential that HPMs of every region in the world have the latest digestible data and understanding to initiate an informed course of action. This urgently requires conducting academic research on several aspects of this highly contagious disease to find effective means of containment and treatment not only for the present but for the future. Tt is critically important for scientists, funding agencies, and HPMs to work together to develop and implement policies that have the greatest likelihood of success in responding to the COVID-19 outbreak. This is particularly challenging in a situation in which much of the evidence is uncertain and evolving rapidly.

As discussed in [36], prior to the COVID-19 pandemic, virology research constituted less than 2% of all biomedical research. However, faced with the pandemic situation, the number of laboratories and investigators that have swiveled to address COVID-19-related research questions is astonishing, likely comprising 10% to 20% of current biomedical investigation, showing the incredible adaptability of the research community. Worldwide financial support rapidly infused COVID-19 research with billions of dollars, leading to a massive influx of publications. More than 300,000 papers have been published since December 2019 in journals of all ranks worldwide [81]. There are also an increasing number of studies being uploaded to preprint servers, such as BioRxiv and MedRxiv, among others. These research articles are continually exploring new observations about the coronavirus and its recently discovered variants from all over the world. Such a collaborative effort is essentially playing a vital role in curbing and controlling the pandemic. However, sometimes a handful number of these research publications might go unnoticed as it is hard to compile these vast knowledge sources even though they are full of potential and could be exploited by the HPMs. At the same time, it has been observed that at times much of the ground-breaking research received significant attention only after a few years of its publication [31, 52]. Rational policy decisions need to combine the latest available scientific evidence — typically published as expert opinions and modeling studies. In addition to HPMs, burgeoning funds for biomedical research indicate the need for understanding global research trends to determine the most pressing priorities. However, in an uncertain and rapidly changing situation such as the COVID-19 pandemic, we cannot assume that the HPMs and funding agencies are able to keep

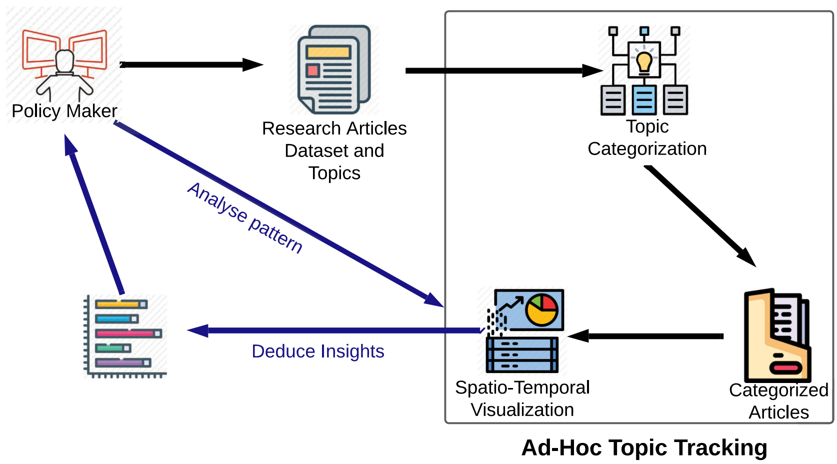


Fig. 1. Flow of our Ad-Hoc Topic Tracking. The Ad-Hoc Topic tracking is shown in a box with two components: topic categorization and spatio-temporal visualization.

pace with the explosion of publications. Therefore, there is a need for an intelligent system that can efficiently track the ongoing research trends around the globe in an ad-hoc way, which can overcome human limits by using the power of artificial intelligence (AI).

Funding agencies and HPMs from different regions have different subjects of interest to scrutinize. Technically, these subjects can be referred to as *topics*. As HPMs are facing unique challenges based on the geographic and demographic profile of each region, their topics of interest are also often different. Topics of interest of HPMs from one region may not directly align with the topics of interest of HPMs from another region; such misalignment of interests can even happen at the personal level. As a consequence, traditional topic modeling-based techniques are unable to track such topics as they are completely unsupervised and do not guarantee alignment with the user's topics of interest provided in an ad-hoc fashion. Therefore, the motivation of this work is to define a new computational task called *Ad-Hoc Topic Tracking* that can accommodate ad-hoc requests. Ad-Hoc Topic Tracking can be defined as a data-mining task in which end users (HPMs /funding agencies) can track topic trends across the globe based on their own topics of interest, which is built upon the following two sequential sub-tasks.

(1) Zero-Shot Topic Categorization: This is a machine learning approach that can predict topics for data it has never seen before (e.g., [72, 73, 84]).
(2) Spatio-temporal Analysis and Visualization: The goal of Spatio-temporal Analysis and Visualization is to visualize and analyze a large dataset, in which data are collected across both space and time. The authors of [6, 9, 20, 25] discuss Spatio-temporal Analysis.

The two sub-tasks are explained in detail in Section 3. A visual illustration of *Ad-Hoc Topic Tracking* is provided in Figure 1 in the context of COVID-19 research tracking. At the very beginning, HPMs gather the research articles, which are input to the Zero-Shot Topic Categorization module. In addition, HPMs provide their topics of interest (in a key-value format). The output of the Zero-Shot Classifier is research articles with topic labels that are used for spatio-temporal visualization. Spatio-temporal visualizations help HPMs in analyzing COVID-19 research dynamics.

We collected data from the COVID-19 Open Research Dataset (CORD-19) [81] and PubTator Central Dataset [82] containing a large number of research papers. Towards achieving the goal of Ad-Hoc Topic Tracking using these datasets, our contributions are as follows:

(1) The PubTator Central Dataset [82] was already curated with six COVID-19-related general research topics: "Disease," "Species,""Chemical,""Cell line,""Mutation,"and "Gene."Thus, we assumed them to be ad-hoc topics of interest without loss of generality and considered them as ground truth for the evaluation of our task.
(2) We proposed, implemented, and evaluated several Zero-Shot Topic Categorization methods (topic-based, embedding-based, and transformer-based) and applied them to categorize 7,000 articles, a subset of research articles (evaluation set) using the ad-hoc topic of interest as labels. Among all the proposed methods, we picked the zero-shot classifier with the best performance on our evaluation set and applied it to further categorize the full set of research articles.
(3) Once research articles were categorized, we performed a spatio-temporal analysis on the labeled articles. We focused on analyzing trends in COVID-19-related research across the globe published in different locations over time by building an interactive spatio-temporal visualization tool, the COVID Research Tracker. For the dataset of this demonstration, we processed around 51,000 articles between December 2019 to December 2021. The COVID Research Tracker consists of three main visualization components:
    (a) **Spatial Visualization:** This component focuses on visualizing COVID-19 articles across different geographic locations in an interactive fashion. The COVID Research Tracker provides a spatial visualization: multilevel granularity-based spatial distribution of COVID-19 research categories through an interactive map interface. For more details, see Section 7.1.
    (b) **Temporal Visualization:** This component presents the interactive time-series visualization of COVID-19 research categories for a particular geographic location. The COVID Research Tracker provides users with options to create visualizations for a particular subset of research categories depending on their interests, allowing them to see those categories' patterns over time. For more details, refer to Section 7.2.
    (c) **Spatio-temporal Visualization:** The third and last component of the COVID Research Tracker allows users to generate visualizations according to their choice and preference to observe the change in research topics across time and space jointly. We provide two types of comparative visualizations in this case: (1) Given two geographic locations, we create dynamic visualizations for the temporal evolution of different research topics; and (2) given two COVID-19 research topics, we create dynamic visualizations to demonstrate their popularity across different geographic locations. More details can be found in Section 7.3.
(4) A demo website has also been implemented, which policy makers can use to create custom spatio-temporal visualizations of COVID-19 research trends: https://bijoy-sust.github.io/Annotation/index.html.

The rest of the article is organized as follows. Section 2 discusses related works, Section 3 describes *Ad-Hoc Topic Tracking*. Section 4 presents the Topic-Categorization procedure. Section 5 provide details of the datasets used; Section 6 presents the experiment results. Section 7 shows a bird's-eye view of the spatio-temporal visualization. Limitations are discussed in Section 8. Our conclusions are presented in Section 9.

2 RELATED WORK

This interdisciplinary work is built upon prior research from multiple areas, including Topic Modeling and Categorization [10, 40, 80], Zero-Shot Learning [78, 83, 84], Spatio-Temporal [14, 54, 65], Data-Driven Decision Support Systems [1, 67, 87], and Policy Making [3, 8, 32]. A discussion on each area and how this work is positioned with respect to the state-of-the-art is as follows.

2.1 Topic Modeling and Categorization

In order to track the global trends of COVID-19 research, it is necessary to first categorize the topics of the research articles. Topic Modeling and Categorization have been studied heavily in the past by the Natural Language Processing (NLP) and Information Retrieval (IR) communities. A summary of existing related works is as follows.

Classical Unsupervised Topic Models: Multiple methods have been proposed in the past for finding latent topics and using them to categorize text [10]. For example, probabilistic topic models perform modestly in identifying topics in unstructured data. The authors of [80] proposed such a topic model, which simultaneously discovers topics and reveals the latent topical structures in text. The authors of [37] proposed a latent-class probabilistic generative model to infer the temporal and spatial patterns of topics that automatically categorizes all points of interest. The authors of [26] proposed a Bayesian model for unsupervised topic segmentation. Karmaker et al. [42] proposed a generative feature-topic model that can mine implicit topics from online reviews through unsupervised statistical learning.

Topic Categorization by Supervised Classification: Multiple previous studies, including [16, 75], have shown that it is possible to categorize topics from well-annotated collections of metadata through supervised learning. The authors of [40] presented a topic model for analyzing and excerpting content-related categories from noisy annotated discrete data such as web pages stored in bookmarks. The authors of [56] combined document classification and topic models using topic modeling to uncover the underlying semantic structure of documents in the collection. The authors of [27] came up with an automatic categorization scheme in which they employed a latent topic model to generate topic distributions given a video and associated text.

Zero-Shot Topic Categorization/Classification: Zero-Shot Topic Categorization has been investigated by researchers in the recent past. In the literature on zero-shot text classification, knowledge of topics is incorporated in the form of word embeddings. The authors of [78] adopted pretrained word embedding for measuring semantic similarity between a label and documents. The authors of [61] attempt to understand how state-of-the-art methods perform on infrequent labels. A few authors developed ***few-shot*** and ***zero-shot*** learning methods for multilabel text classification. The authors of [84] benchmark the Zero-Shot Text Classification problem by providing unified datasets, standardized evaluations, and state-of-the-art baselines. The authors of [63] published 2 suitable datasets for the Zero-Shot Text Classification task. The authors of [83] studied the zero-shot intent detection problem, which aims to detect emerging user intents without any labeled utterances. The authors of [88] incorporated four kinds of semantic knowledge — word embeddings, class descriptions, class hierarchy, and a general knowledge graph —into their proposed framework to deal with Zero-Shot Text Classification. Pushp and Srivastava [58] implemented the "TRAIN ONCE, TEST ANYWHERE"approach, which involves a training model to tackle unseen sentences, tags, and even new datasets provided. Puri and Catanzaro [57] proposed generative models for zero-shot text classification. Another line of researchers [86] characterized the performance of discriminative and generative long short-term memory (LSTM) models for text classification and confirmed in a series of experiments in zero-shot learning settings that generative models substantially outperform discriminative models. The authors of [18] proposed a new model that combines BERT with Label-Wise Attention Networks (LWANs) and showed that the new model improved ***few-shot*** and ***zero-shot*** learning. Researchers also implemented Zero-Shot Text Classification via Knowledge Graph Embedding for Social Media Data in [21].

Topic Categorization Tools: Knowtator [53] is a general-purpose text categorization tool that facilitates the manual creation of training and evaluation corpora for a variety of bio-NLP tasks. The authors of [12] developed GATE Teamware, which is an open-source, web-based, collaborative text categorization and annotation framework. Seeker is a platform for large-scale text analytics developed by the authors of [24], and SemTag is an application written on top of the platform to perform automated semantic tagging/categorization of large corpora.

2.2 Research on COVID-19–Pandemic Related Policies

Policy making during a pandemic can be extremely challenging. As COVID-19 is a new infection and its global impacts are unprecedented, decisions are taken in a highly uncertain, complex, and rapidly changing environment. To tackle this challenge, a line of work focused on studying different aspects of policy making and their implications. The authors of [4] discussed what policy makers need to know about COVID-19 protective immunity. The authors of [35, 64, 76] discussed crushing and curbing COVID-19 with the help of well-informed policy making. The authors of [23, 39] presented a few considerations for conservation policy makers to support and rethink the development of impactful and effective policies in light of the COVID-19 pandemic. The authors of [30] identified the key factors that school leaders have had to react and respond to when creating policy in the context of COVID-19. The authors of [19] described the impact of COVID-19 on children, offering some outcome-based suggestions to policy makers and caregivers to mitigate the negative impact of the pandemic on COVID-affected families and children. Several studies have been performed, including those described in [3, 8, 32] for modeling the perspective or decision-making of the parents, clinicians, policymakers, and others. As we know, scientific insights are important factors for policy makers and decision-making. The authors of [85] discussed the co-evolution of policy and science during the pandemic and, in [51], the authors have described human–computer interaction interventions for science communication.

2.3 Zero-Shot Learning (ZSL) on COVID-19 Related Datasets

Ever since the beginning of the pandemic, a large number of articles, clinical notes, and reports have been made available for various research. Studies in [44, 47, 71] used Zero-Shot Learning (ZSL) for classifying such biomedical articles and clinical notes. Whereas the authors of [48] applied ZSL for a COVID-19 literature search, the authors of [60] implemented ZSL for object detection in medical imaging on a COVID-19 Chest X-Ray (CXR) dataset. To stop the spread of COVID-19, people were asked to wear masks that, in turn, resulted in a large number of fatalities and safety concerns. For that reason, a group of researchers [69] employed ZSL for face detection by identifying the face mask. The sudden spread of the global pandemic COVID-19 had led to panic, speculations, and the spread of misinformation. A line of researchers [41, 74] focused on identifying fake tweets related to COVID-19 using ZSL. The authors of [5], collected and published the COVID-19 utterance dataset and made an attempt at cross-lingual transfer learning for intent detection using ZSL.

2.4 Data-Driven Decision Support for COVID-19 Pandemic Management

Due to the increasing popularity of Data Science and AI, researchers have spent a lot of effort to provide data-driven decision support for COVID-19 pandemic management. We discuss some of these efforts briefly as follows.

Spatio-temporal Visualization: Data visualization for COVID-19 symptoms, spread and prediction, demographic data analysis, and enhancing awareness have been major focuses for research in the visualization domain. Some notable works include [6, 9, 20, 25, 62]. Spatio-temporal analysis for exploring the effect of temperature and other environmental correlations on COVID-19 in Spain is presented by the authors of [14, 54]. The authors of [65] performed spatio-temporal analysis for finding medical resource deficiencies in the United States during the COVID-19 pandemic, while the authors of [66] performed spatio-temporal analysis for hot-spot detection. Another group of researchers conducted spatio-temporal analysis on various topics, such as government support, cases of new infections, deaths, and panic buying [7, 46].

Research Trend Analysis: A school of researchers worked on analyzing research trends in various areas after the emergence of COVID. Examples include global research trends in COVID-19 vaccines [1]; identifying research trends and gaps in the context of COVID-19 [67, 87]; and investigating the emerging COVID-19 research trends in the fields of business, management, and marketing science [49, 79].

Predictive Modeling: Another group of researchers explored AI-based predictive modeling techniques for COVID-19 and similar pandemic-related crises. For example, the authors of [38, 70, 77] aimed at employing AI solutions to analyze and prepare us for prevention of and fighting COVID-19 and other pandemics. The authors of [33] focused on AI-enabled COVID-19 outbreak analysis and prediction. The authors of [28] aimed to develop AI-based methods to quantify disease severity and predict COVID-19 patient outcomes.

Sentiment Analysis: With the isolation of people from physical public spaces in response to the COVID-19 pandemic, online platforms have become even more prominent tools to analyze public opinion and concerns (such as social network discussion). Individuals, organizations, and governments are using social media to communicate with each other on a number of issues relating to the pandemic. A school of researchers has performed sentiment analysis on Twitter posts in order to understand public emotions [34, 43, 45]. Another line of research work analyzed conspiracy theories propagated over Twitter [2, 15].

Fake News Detection: Since the spread of the coronavirus disease, uninhibited misinformation is also spreading over traditional and social media at a rapid pace. A few attempts have been made to identify fake/incorrect information in order to prevent it from spreading [13, 43, 68].

2.5 Contribution of Our Work

In contrast to the existing research, our work is more interdisciplinary, with the goal of helping policy makers with AI-assisted, never seen before topics of interest in an ad-hoc fashion. To achieve this challenging goal, we introduce a new computational task called *Ad-Hoc Topic Tracking*, which can be divided into a sequence of two sub-tasks: (1) Zero-Shot Topic Categorization and (2) Spatio-temporal Analysis and Visualization. This powerful combination enables us to serve multiple policy makers with different preferences as well as different topics of interest in an ad-hoc fashion. In the end, we also present how topic categorization combined with spatio-temporal analysis can be leveraged to create interactive visualization that can greatly help HPMs and funding agencies with better decision-making.

3 WHAT IS AD-HOC TOPIC TRACKING? HOW DOES THAT HELP POLICY MAKERS?

Predominantly due to local context and sociopolitical influence, policy makers and funding agencies of different regions have different goals to focus on. Thus, a system that deals with only a predefined set of topics cannot accommodate requests from various end users (policy makers in this case) and is unsuited for real-world applications. In contrast, an ad-hoc system that can serve users of disparate interests is more useful. In our study, *Ad-Hoc Topic Tracking* can be defined as a system in which end users such as policy makers and funding agencies can track research trends across the globe based on their own topics of interest. The *Ad-Hoc Topic Tracking* task can be imagined as a 2-step process: (1) Zero-shot Topic Categorization and (2) Spatio-temporal Visualization.

(1) Zero-Shot Topic Categorization: This method has gained much popularity recently. In Section 2, we have discussed the existing works on *Zero-Shot Topic Categorization*. The motivation behind choosing the zero-shot approach is the following: while analyzing a large corpus of documents, we first need to structure them, that is, the first step is to categorize all documents in the

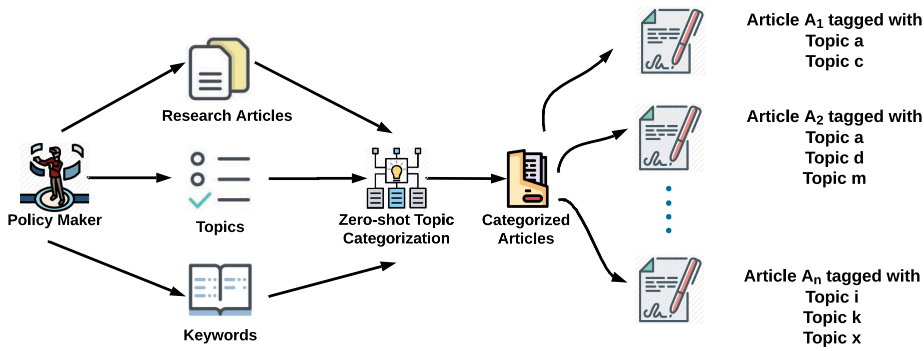


Fig. 2. The zero-shot topic categorization. The health policy maker (domain expert) provides a set of research articles, topics, and topic-related keywords. The categorization algorithm then uses an unsupervised approach to assign topics to each document.

corpus with topic-related metadata. However, instead of a set of predefined topics, the definition of topics in our problem setup comes from the users in real time (ad-hoc basis), who have the best knowledge of the application scenario (such as policy akers). Therefore, we adopted a zero-shot approach for topic categorization for this work, which provides the end user with the control to maximize the utility of the outcome of the categorization process.

In order to highlight the difference between traditional topic categorization and its zero-shot counterpart, we compare their formal definitions.

Definition 1 (Traditional Topic Categorization). Given a collection of documents D and a set of **predefined** topics T, categorize each document $d \in D$ with one or more topics in T.

Thanks to the **predefined** set of topics T, the traditional *Categorization* task can benefit from fine-tuning based on a carefully designed training set for supervised learning.

Definition 2 (Zero-Shot Topic Categorization). Given a collection of documents $D = \{d_1, d_2, \ldots, d_n\}$, a user x and a set of **user-defined** topics $T = \{t_1, t_2, \ldots, t_m\}$ provided in **real-time**, categorize each document $d_i \in D$ with zero or more topics from T **without any further fine-tuning**.

Note that it is possible that two different users will provide a different set of topic definitions for the same dataset based on their application needs and end goals, which is totally acceptable in our problem setup. This essentially means that creating customized training datasets beforehand for fine-tuning is no longer possible because the target topics are provided in real time. This essentially makes *Zero-Shot Topic Categorization* an unsupervised task. We also assume that each topic t_j is expressed as a word/phrase and that the user has the option to provide a list of additional keywords/key phrases K_j associated with each topic $t_j \in T$. Our zero-shot problem setting assumes that the end user provides all of the documents, the target topics, and a topic-keyword dictionary (optional) as inputs in real time. The user here is usually a domain expert (e.g., policy maker or funding agency) with specialized knowledge or skills in a particular area of endeavor.

For a better demonstration of the zero-shot topic categorization framework, we present an intuitive example in Figure 2. Consider the domain expert (i.e., policy maker) who is analyzing a large volume of research articles and wants to computationally categorize the articles with research-related topics such as “Gene,” “Cell line,” and “Species.” For this real-life use case, the domain expert will provide the collection of documents (to be categorized) as well as a set of topics to be used as tags for categorizing the documents. Additionally, the domain expert may also provide a list of relevant keywords associated with each topic that can be used as expert guidance for the categorization process. The zero-shot topic categorization algorithm then labels each document by associating it with relevant topics.

As we mentioned, the user may also provide a list of relevant keywords/clues associated with each topic to guide the categorization process. For example, if a medical professional wants to categorize some documents with "Heart Health," some helpful keywords may be "Stroke," "Cardiovascular," and "Hypertension." Section 6.2 provides a discussion on why these keywords are particularly helpful for the *Zero-Shot Topic Categorization* task. Further, a topic t_j may not occur by its name/phrase *explicitly* in a document d_i. For example, a document about "Chemical Products" may not include the exact word "Chemical Products" but still talk about "ethanol," "chloroquine," "heparin," and "drugs." Thus, the topic "Chemical Products" is implicit in this document; it is equally important to identify the implicit topics within a document as well as the explicit topics. Although the user-provided optional keywords may help mitigate this issue to some extent, it is almost impossible to provide a comprehensive list of keywords that can capture all possible ways the "Chemical Products" topic can be mentioned. At the same time, a single appearance of a keyword may not always mean that the document as a whole is focused on the corresponding topic. To summarize, neither the presence nor absence of keywords are sufficient to infer the correct topics associated with a document; they are simply informative clues from the user end.

(2) Spatio-temporal Visualization: After the zero-shot method categorizes the corpus with the user-provided topics of interest, a spatio-temporal visualization is created by jointly analyzing the topic-related metadata along with time and location information. Text, time, and location information can be jointly exploited beyond simple statistics, such as finding frequent patterns, spatial changes, outliers, and spatio-temporal clusters. Spatio-temporal visualization shows the changes in information in space and time. It has the natural advantage of revealing overall trends and movement patterns; thus, it constitutes an important instrument in terms of decision-making. Further discussion on this can be found in Section 7.

4 ZERO-SHOT TOPIC CATEGORIZATION METHODS

As discussed in Section 3, *Zero-Shot Topic Categorization* is a special type of machine learning task in which users can define their own topics of interest as labels and then run a classifier to assign a probability to each label. Zero-shot learning is about leveraging pretrained supervised models without any training data available for fine-tuning. Classification is performed by associating observed data and user-defined labels through some form of auxiliary information (additional clues from the user) with the help of pretrained models (trained on a generic corpus). Therefore, the inputs to the Zero-Shot Topic Categorization method are the following:

(1) Corpus of documents to be categorized
(2) Labels, that is, a list of topics of interest (comes from the user in real time)
(3) Auxiliary information (e.g., textual descriptions/keywords related to each topic of interest)

Given these inputs, we discuss a wide variety of zero-shot topic categorization approaches in the next subsection.

4.1 Topic Modeling Based Zero-Shot Approaches

As classical topic models are completely unsupervised and, hence, cannot be directly used to perform classification/categorization, we extended the classical topic models for zero-shot topic categorization by associating the topic distribution of a document with the topic distribution of auxiliary information of the target labels (textual descriptions/keywords related to each topic of interest). We discuss two such approaches here.

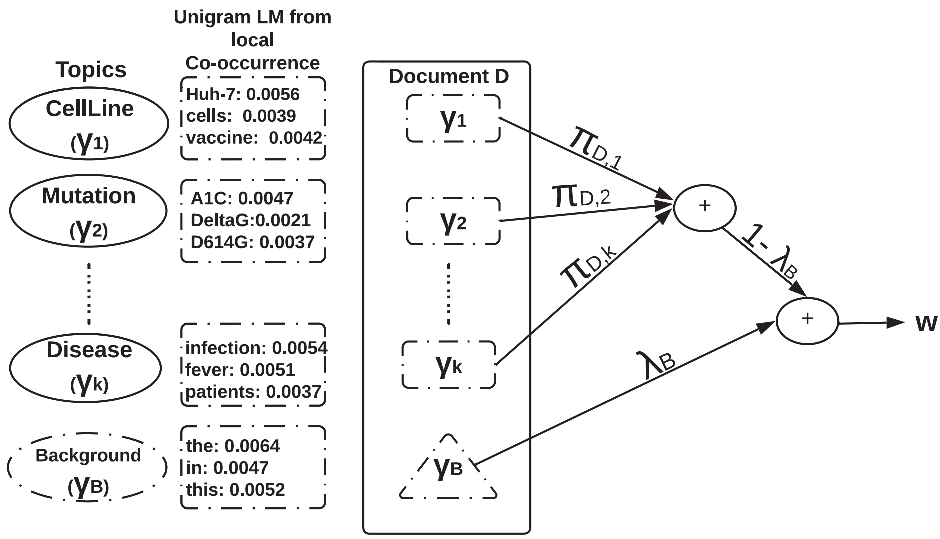


Fig. 3. Generative process for the generative feature language model with high confidence on auxiliary information.

4.1.1 Generative Feature Language Model with High Confidence on Auxiliary Information (GFLM). As our first zero-shot baseline, we used the Generative Feature Language Model proposed by the authors of [42]. The article introduced a zero-shot topic categorization technique that can mine the implicit mentions of topics effectively through unsupervised statistical learning. The GFLM is shown in Figure 3. The steps of the GFLM can be summarized as follows.

(1) Each sentence is created by generating each of the words in the sentence independently.
(2) To generate a word w in sentence S, we first decide whether we would generate the word using the background model γ_B or a topic language model γ_i. We make this choice according to $\lambda_B \in [0, 1]$, which is a parameter indicating the probability of using the background model instead of a topic language model. Thus, the probability of choosing a topic language model would be $1 - \lambda_B$.
(3) If we have chosen the background language model, we would sample the word from the distribution $p(w|\gamma_B)$; otherwise, we would further make a decision on which of the k topic language models to use. This decision is made based on another set of parameters $\{\pi_{D,i}\}$, where $i = 1, \ldots, k$, and $\sum_{i=1}^{k} \pi_i = 1$. $\pi_{D,i}$ is the probability of choosing topic language model γ_i to generate the word. Thus, with probability $\pi_{D,i}$, we would sample the word using $p(w|\gamma_i)$.
(4) This process would be repeated to generate all of the words in a sentence, and all of the sentences would be generated in the same way, each sentence being generated using a set of sentence-specific topic choice parameters $\pi_{D,i}$.

According to this generative model, the probability of observing a word w in a document D assuming k different topics is the following:

$$P_D(w) = \lambda_B P(w|\gamma_B) + (1 - \lambda_B) \sum_{i=1}^{k} \pi_{D,i} P(w|\gamma_i), \tag{1}$$

where, λ_B is the proportion of words generated from a background language model (mostly stop words and common words), $P(w|\gamma_B)$ is the unigram distribution of the background language model,

π denotes the topic distribution, and $P(w|\gamma_i)$ represents the unigram distribution associated with each topic. Thus, the log-likelihood of observing the entire set of documents C from a mixture model with unknown parameters Λ is

$$\log P(R|\Lambda) = \sum_{D\in C}\sum_{w\in V}[c(w,D)\times \log\{\lambda_B P(w|\gamma_B) + (1-\lambda_B)\sum_{i=1}^{k}(\pi_{D,i}P(w|\gamma_i))\}]. \tag{2}$$

The authors of [42] put high confidence in the auxiliary information provided for each topic. They achieved this by precomputing $P(w|\gamma_i)$ based on explicit mentions of the topic words (simple string matching followed by computing the relative frequency of neighboring words) and never changing/revising it afterward. Assuming λ_B and γ_B to be known, the $\pi_{D,i}$ parameters (topic distributions) of this generative model are optimized by maximizing the log-likelihood of the data (Equation (2)). This estimation is performed using an Expectation-Maximization algorithm. We implemented the following E-step (Equations (3) and (4)) and M-step (Equation (5)):

$$\textbf{E Step:}\qquad P(z_{D,w}=t) = \frac{\pi_{D,f}^{(n)}P(w|\gamma_t)}{\sum_{t'=1}^{k}\pi_{D,t'}^{(n)}P(w|\gamma_{t'})} \tag{3}$$

$$P(z_{D,w}=B) = \frac{\lambda_B P(w|\gamma_B)}{\lambda_B P(w|\gamma_B) + (1-\lambda_B)\sum_{t'=1}^{k}\pi_{D,t}^{(n)}P(w|\gamma_t)} \tag{4}$$

Here, $P(z_{D,w} = f)$ indicates the probability that word w in document D is generated from feature topic γ_f given that w is not generated from the background model γ_B.

$$\textbf{M Step:}\qquad \pi_{D,t}^{(n+1)} = \frac{\sum_{w\in V}c(w,D)(1-P(z_{D,w}=B))P(z_{D,w}=t)}{\sum_{t'=1}^{k}\sum_{w\in V}c(w',D)(1-P(z_{D,w'}=B))P(z_{D,w'}=t')} \tag{5}$$

Here, $c(w,D)$ denotes the count of word w in document D. Interestingly, A key component for re-estimating π (topic distributions) is $c(w,d)(1-p(z_{D,w}=B))p(z_{D,w}=t)$, which can be interpreted as the allocated counts of w to topic t. Intuitively, we use the inferred distribution of z values from the E-step to split the counts of w among all of the topics. The number of fractional counts of w that t can get is determined based on the inferred likelihood that w is generated by topic t. Once we have such a fractional count of each word for each topic, we can easily pool these split counts to re-estimate π.

After the EM algorithm converges, one knows the identities of each word $P(z_{D,w} = t)$ and $P(z_{D,w} = B)$, that is, the degree to which the background model or some specific topic contributed to the generation of a particular word. One also knows the topic distributions $\pi_{D,t}$, that is, to what proportion a particular document D is generated from some topic-of-interest t. Based on these quantities, topic distributions within various documents can be inferred in two different ways, which was called **GFLM-Word** (GFLM-W) and **GFLM-Sentence** (GFLM-S), respectively.

In the case of GFLM-Word, given a document D, it looks at each word w and adds a topic t to the inferred topic list if and only if $p(z_{D,w}=t)\times(1-p(z_{D,w}=B))$ is greater than some threshold θ for at least one word in D. The philosophy behind this formula is that if any particular word w has a small probability of being generated by a background model but has a higher probability of being generated from some topic t, then word w is likely referring to topic t. Here, the decision is made solely by looking at individual words, not the entire document.

In the case of GFLM-Sentence, given a document D, it looks at the contribution of each topic t in the generation of the sentence, that is, $\pi_{D,t}$, and infers t^* as the topic only if π_{D,t^*} is greater than some user-defined threshold θ. Here, the decision is made at the sentence level, not at the word level.

4.1.2 Generative Feature Language Model with Moderate Confidence on Auxiliary Information. One particular limitation of the Generative Feature Language Model (GFLM) proposed b the authors of [42] is the high confidence it puts in auxiliary information. What if the auxiliary information is incorrect/noisy? To address this limitation, we further extended the GFLM where topic distributions are initialized with the help of auxiliary information; however, topic distributions are revised as the E-M estimation process continues iterations, allowing more flexibility. Mathematically, Equation (6) is included in the M-step, which further adjusts the topic distributions. This technique can be viewed as putting moderate confidence in auxiliary information, which is more realistic.

$$\textbf{M-Step:} \qquad P^{(n+1)}(w|\gamma_t) = \frac{\sum_{D \in C} c(w, D)(1 - p(z_{D,w} = B))p(z_{D,w} = t)}{\sum_{w' \in V} \sum_{D \in C} c(w', D)(1 - p(z_{D,w'} = B))p(z_{D,w'} = t).} \tag{6}$$

For re-estimating $p(w|\gamma_t)$, the key term is $c(w, D)(1 - p(z_{D,w} = B))p(z_{D,w} = t)$, which can be interpreted as the allocated counts of w to topic t. Intuitively, we use the inferred distribution of z values from the E-step to split the counts of w among all topics. Again, the number of split counts of w that t can get is determined based on the inferred likelihood that w is generated by topic t. Once we have such a split count of each word for each topic, we gather the split counts of a word w toward topic t from all documents in the collection and then normalize these counts among all of the words in all of the documents to re-estimate $p(w|\gamma_t)$. Finally, topic distributions within various documents can be inferred in two different ways in a similar fashion as *GFLM-Word* and *GFLM-Sentence*. We call these two approaches **GFLM-Word-Moderate (GFLM-W-M)** and **GFLM-Sentence-Moderate (GFLM-S-M)**, respectively.

4.2 Classical Word-Embedding Based Zero-Shot Approaches

Classical word embeddings are a popular way to encode text data into a dense real-valued vector representation. In order to implement a zero-shot classifier, we encoded both the input document and the target topics using pretrained word embeddings. Then, we computed vector similarity between the input document encoding and each target topic encoding separately. The details of the classical word-embedding-based zero-shot approach are provided here.

(1) The inputs — article text, topics, and auxiliary information (keywords) — are provided by the user.
(2) For each topic t, encode it by averaging the pretrained embeddings (e.g., Glove, Word2Vec) of each word present in the auxiliary information of the corresponding topic. For example, if the target topic was "Chemical" and auxiliary information was provided as a list of keywords/clue-words — "oxygen," "hydroxychloroquine," "chloroquine," "remdesivir," "creatinine," "oseltamivir," etc. — then we compute the topic embedding by taking an average of all of the embeddings of all of these words.
(3) Articles are represented in 2 different ways. (a) *Average Sentence Level Embedding:* For each input article D, we encode the document by averaging the pretrained embeddings (e.g., Glove, Word2Vec) of each word present in that document. (b) *Dictionary of Word Embeddings*: Extract word embedding of all of the words in an article and instead of taking the average, we save them individually as a key-value pair.
(4) Once we obtain the topic embeddings and document embeddings, the next step is to assess the semantic similarity between these two embeddings. For semantic similarity, we used 2 different metrics: (a) Euclidean distance and (b) cosine similarity. Since we have 2 types of embeddings for each article (word and sentence level), the inference of topics is performed separately as well. For the sentence-level embedding, *Cosine Similarity* or *Euclidean Distance* is measured between the sentence embedding and topic embedding (computed in step 2).

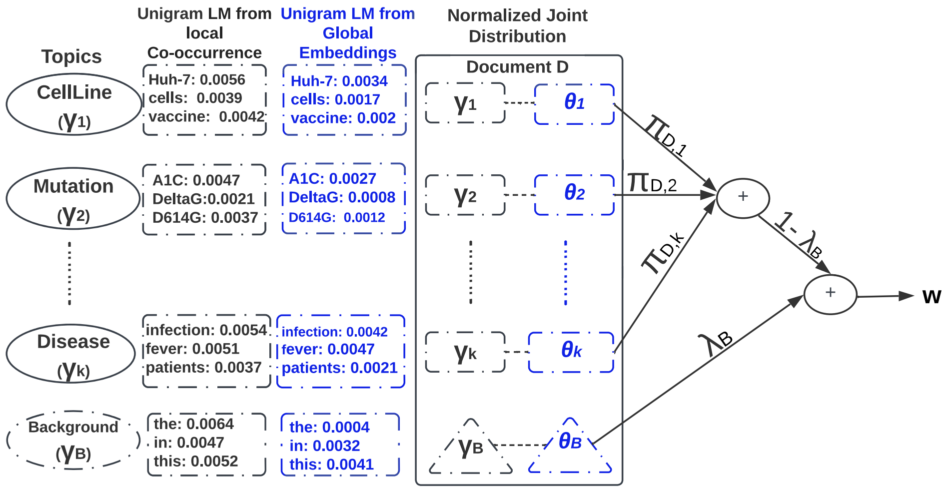


Fig. 4. Generative process for Joint Topic and Embedding-Based Zero-Shot Approach. The blue portion denotes the embedding-based topic-word distribution and embedding-based background-word distribution.

For word embedding, distance (Cosine/Euclidean) is measured between each individual word embedding and topic embedding. Based on the similarity/distance and a user-defined threshold (theta), the zero-shot categorization method assigns a topic to the input article if the similarity is higher than the user-defined threshold.

(5) Finally, the output of the zero-shot classifier is the set of input documents and the corresponding inferred topics associated with them. In summary, we have implemented four classical embedding-based zero-shot topic categorization techniques.
- Sentence Embedding Euclidean Distance (Euclidean-Sentence)
- Word Embedding Euclidean Distance (Euclidean-Word)
- Sentence Embedding Cosine Distance (Cosine-Sentence)
- Word Embedding Cosine Distance (Cosine-Word)

4.3 Joint Topic and Embedding-Based Zero-Shot Approaches (Hybrid)

This technique is an extension of the GFLM approach described in Section 4.1.2. We call it the "Joint Topic and Embedding-Based Zero-shot Approach." Here, we revised the E-step from Equations (3) and (4) by including word embeddings that can further strengthen the zero-shot classifier by injecting external knowledge learned by these embeddings. The generative process is presented in Figure 4. The process is the same as for the GFLM, except that in the Hybrid approach, words are sampled from a joint distribution (instead of a single unigram language model) of one local distribution based on co-occurrences in the input documents and a global distribution based on embeddings learned from a global large corpus, i.e., ($P(w|\theta_t)$). The benefit of this approach is to leverage external knowledge learned from large text corpora to improve the topic inference process. In a similar spirit, an Embedding-Based Background-Word Distribution, i.e., ($P(w|\theta_B)$) is introduced (shown in the blue box).

Initially, we experimented with pretrained word embeddings such Stanford's Word2vec [50], GloVe [55], and FastText [11]. We observed that these pretrained embeddings dictionaries are missing many words related to COVID research, e.g., "SARS-CoV-2," "interleukin-8," "miR-93-5p," and more. Hence, to improve the embeddings, we trained word embeddings on our large corpus of COVID-19 research articles and used the custom word embeddings in the zero-shot classification task. For the Joint Topic and Embedding-Based Zero-Shot Approach, the revised formulas for E-step are shown here, while the M-step remains identical to the original GFLM-Moderate approach.

$$P(z_{D,w} = t) = \frac{\pi_{D,t}^{(n)} P(w|\gamma_t) P(w|\theta_t)}{\sum_{t'=1}^{k} \pi_{D,t'}^{(n)} P(w|\gamma_{t'}) P(w|\theta_{t'})} \tag{7}$$

$$P(z_{D,w} = B) = \frac{\lambda_B P(w|\gamma_B) P(w|\theta_B)}{\lambda_B P(w|\gamma_B) P(w|\theta_B) + (1 - \lambda_B) \sum_{t'=1}^{k} \pi_{D,t'}^{(n)} P(w|y_t') P(w|\theta_t')} \tag{8}$$

The E-step formula (Equations (7) and (8)) introduced here include two new terms: $P(w|\theta_t)$ and $P(w|\theta_B)$. We define these terms as Embedding-Based Topic-Word Distribution, ($P(w|\theta_t)$), and Embedding-Based Background-Word Distribution, ($P(w|\theta_B)$). The following are detailed steps that show how we obtained these two distributions.

(1) As we know, a topic is essentially a probability distribution over all of the words from the vocabulary. For example, the topic "Chemical" may contain different words such as "oxygen," "hydroxychloroquine," "chloroquine," and "remdesivir." To generate the Topic Embedding, we first picked the top 30 high-probability words from each topic model $P(w|\gamma_i)$ and averaged their word embedding to denote the topic embedding.
(2) Once we calculate the topic embedding, we measure the distance between the embedding of each word in the article and the topic embedding. For distance measures, we used Euclidean and cosine distances.
(3) The distance computed in the previous step (step 2) is then multiplied with Topic-Word Distributions $p(w|\gamma_t)$ and normalized to derive the Embedding-based Topic-Word Distribution, ($P(w|\theta_t)$).
(4) The same steps have been followed to generate the Embedding-Based Background-Word Distribution, ($P(w|\theta_B)$). The only difference is that instead of the topic model, we picked the top 30 high-probability words from the background language model, $P(w|\gamma_B)$ (refer to step 1).

Note that there has been no change in M-Step Equations (5) and (6). Topic Coverage (π) and Topic-Word Distributions $p(w|\gamma_t)$ are computed similarly as we did in the Generative Feature Language Model with Moderate Confidence on Auxiliary Information approach. The whole purpose of this joint model is to improve topic models' performance by augmenting the power of word embeddings.

4.4 Contextual Embedding based Zero-shot Models

We implemented recent contextual embedding-based zero-shot models. We specifically focused on contextual embeddings because they assign each word a unique representation based on its context. Therefore, they can capture the nuances of meanings across varied contexts. Contextual embeddings encode knowledge in a way that can be transferred across different domains. In this work, we used transformers (BERT and BART) and Bi-LSTM Language models (ELMo) based on contextual embedding techniques for zero-shot categorization of research articles. The overall process of zero-shot categorization using sentence embeddings is similar to that of classic word embeddings, as demonstrated. The main difference is in step 3, in which, as an Embedding Generator, we used state-of-the-art contextual word embeddings like BERT, BART, and ELMo. The rest of the steps work as before.

4.5 Sentence Embedding-Based Zero-Shot (ZS) Models

Sentence embedding attempts to encode a sentence or short paragraphs into a fixed-length vector (dense vector space). Then, the vector is used to represent that sentence/paragraph for a subsequent downstream task [17, 59]. In contrast to word embeddings, sentence-level representation models (i.e., sentence embeddings) map the full (whole) sentence to an embedding representation to capture the overall semantics more accurately. Such context-sensitive sentence embeddings have proven to improve various downstream tasks in many domains. Therefore, in addition to word embeddings, we used several sentence embedding techniques for performing zero-shot topic categorization of research articles.

We explored three sentence embedding techniques: Universal Sentence Encoder (USE) [17], InferSent [22], and SBERT [59]. The benefit of sentence embedding over word embedding is that, in the case of word embedding, individual word embeddings are averaged to derive a sentence/paragraph level representation, which often destroys the semantic consistency at the sentence/paragraph level. However, sentence encoders encode the meaning of the whole sentence in a single vector as they are trained specifically to do that without any requirements of averaging. It has been demonstrated in the literature that sentence encoders can capture the semantics better at the sentence level compared with individual word embeddings.

The overall process of zero-shot categorization using sentence embeddings is similar to that of classic and contextual word embeddings. The main difference is in step 2, in which, as an Embedding Generator, we used state-of-the-art sentence encoders such as USE, InferSent, and SBERT. The rest of the steps work as before.

5 DATASETS

In response to the COVID-19 pandemic, the White House and a coalition of leading research groups have prepared the COVID-19 Open Research Dataset (CORD-19) [81]. CORD-19 is a resource of over 300,000 full-text scholarly articles.[1] We extracted specific information from each of the articles, including the abstract, first author location, title, and much more. We further searched other available datasets over the Internet and found a relevant database called PubTator [82]. All records from both datasets span the time frame of January 2020 to December 2021. A brief description of both datasets is presented here.

CORD-19: This dataset is the largest collection of coronavirus literature. It is updated regularly as new research is published in peer-reviewed publications and archival services such as bioRxiv, medRxiv, and others. The dataset directory contains many files, among which we used **`Pdf_Json`**(e.g., .json files) and **Metadata** (e.g., .csv files) files to build our final dataset named Kaggle CORD 19. **`Pdf_Json`** contains the author's name, location, and, most significantly, the full text of COVID-19 research papers. **Metadata** is a collection of PMC ID, Source, Publication Time, DOI, and so on, which are integral parts of any scientific literature. Here, the PMC ID helps us to build our final gold standard dataset by serving as the primary key.

PubTator Central: The PubTator Central [82] dataset is a collection of PMC ID (paper unique ID) and full-text articles with categorization of biomedical topics such as "Gene," "Mutation,""CellLine,""Species,""Disease,"and "Chemical"for each paragraph in the corresponding articles. Each article is labeled with one to many topics; in Figure 5, we have shown the overlap in mappings between topics for 2 different articles. The diagram shows that both articles are related to the topic "Disease"; thus, there is indeed some overlap. Article 1 is also related to "Chemical" and "Species," whereas Article 2 is labeled with "CellLine" and "Gene." Since none of them are related

[1] This public dataset is available at https://www.kaggle.com/allen-institute-for-ai/CORD-19-research-challenge.

to the topic "Mutation," the topic belongs to the area outside of the circle. For our experiments, we assumed these topics to be user-defined ad-hoc topics for this work, which essentially provided us with ground truths to evaluate our *Zero-Shot Topic Categorization* methods quantitatively.

5.1 Data Preprocessing

After fetching all of the data from the *CORD-19* collection, we found that some of the location's name has been either misspelled or informally written. We preprocessed the incorrect location entries into a standard format using GeoPy API[2]. The preprocessed data dictionary was then cleaned and later merged with the metadata file to recreate the final dataset, called Kaggle CORD 19.

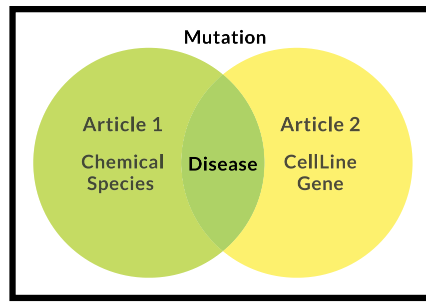


Fig. 5. Mapping among six topics for the two different articles.

For the *PubTator Central* database [82], the structure of each record is as follows: each record contains the full text of the paper and the full text is an aggregation of individual paragraphs. We iterated through each paragraph of the full text and extracted the corresponding category along with the paragraph text. After the extraction process, these data were merged with Kaggle CORD 19 using the PMC ID. Finally, records containing missing attributes/null values were removed, resulting in a total of 51K records. Each record belongs to a particular research article along with the ground-truth topic labels extracted from PubTator. This constitutes our final dataset, called Gold Standard (Covid) Dataset[3]. Attributes of this gold standard dataset are presented here.

- PMC ID: The paper-unique ID or MEDLINE identifier of a manuscript
- Title: The title of the manuscript
- Body: The full text of the manuscript
- Publication date: The date of publication of manuscript
- Author's Location: First Author's location
- Topics: The research topics, such as "Gene,""Mutation,""CellLine,""Species,""Disease,"and "Chemical."

Due to computational resource constraints, we further sampled 7K records from 51K records to perform our initial exploration and model selection tasks for topic categorization. Table 1 provides an overview of this dataset. However, our spatio-temporal plots were created on the whole dataset (51K records) once we settled on the ultimate zero-shot topic categorizer.

5.2 Prevalence of Implicit Mentions

To establish the simplest baseline, we first checked whether the ground-truth topics can be identified by simply matching the topic name/topic phrase against the article text, the results of which are reported in Table 2. For each article, topics inferred by simple string matching were compared against the ground-truth topics to compute the True Positive, False Positive, and False Negative statistics, which are defined next.

[2] https://geopy.readthedocs.io/en/stable/.
[3] The final data-set is available at: https://drive.google.com/drive/folders/1Gv1Xfd76txNrDBBAIqxiIZCyy1kYiDyb?usp=sharing.

Table 1. An Overview of the Gold Standard Dataset

Dataset ->	**COVID Dataset**
Total # of Articles	7,001
# of Topics	6
Avg. # of Topics per article with ≥1 topic	2.692

Table 2. Categorization Result Based on String Matching by Topic Name

Topic Name	**Total Count**	**True Positive**	**False Negative**	**False Positive**
CellLine	430	0	430	0
Chemical	2,908	418	2,490	239
Disease	6,860	895	5,965	94
Gene	2,710	684	2,026	2,856
Mutation	130	75	55	374
Species	5,803	454	5,349	35

- **True Positive:** Number of topics correctly extracted.
- **False Negative:** Number of actual topics not extracted.
- **False Positive:** Number of topics incorrectly extracted.

A closer look into the initial results (False Negative values in Table 2) and our dataset revealed that the dataset is composed of lengthy research articles (approximately 1,981 words per article), and each article is a complex representation of various topics, entities, and events. As expected, simply checking the topic name in the text did not yield a high-quality automatic categorization. We also found that most of the topics are not explicitly mentioned in the article; thus, they are "Implicit Mentions." The difference between *explicit* and *implicit* mentions can be further clarified through an example. We consider a topics to be *explicit* if the topic's name/topic phrase is explicitly mentioned in the article text. For example, the following sentence is from an article related to the topics ***Disease*** and ***Species***, ***"SARS-CoV-2 escape from a highly neutralizing COVID-19 convalescent plasma,"*** mention of the word ***COVID-19*** somewhat describes the disease but the topic ***Species*** denoted by the keyword ***SARS-CoV-2*** is implicit. Implicit topics are those for which the topic name is not directly mentioned in the article text; rather, the topic is implied through the text. For example, the text ***"Remdesivir Efficacy in COVID-19 Treatment: A Randomized Controlled Trial"*** does not contain the word ***Chemical***, yet when a domain expert observes the word "Remdesivir," the expert can easily relate it to the ***Chemical*** topic. We consider these cases as implicit mentions of the target topic. Based on the above observation, we performed a detailed analysis on explicit and implicit topics for the gold standard dataset, the results of which are presented in Table 3.

6 EXPERIMENTS FOR ZERO-SHOT TOPIC CATEGORIZATION

6.1 Performance Measures

To measure the performance of each zero-shot categorization approach, we use three popular metrics available in the literature: Precision, Recall, and the F_1 score. For each article, the model inferred topic(s) were compared against the list of "gold" topic(s) to compute the true-positive, false-positive, and false-negative statistics for that article. Then, all such statistics for all of the articles in a dataset were aggregated and used to compute the final Precision, Recall, and F_1 score.

Table 3. Analysis of Explicit and Implicit Mentions of Gold Standard Dataset

Statistics	COVID Dataset
Total explicit mentions	8,683
Total implicit mentions	10,158
Percentage of explicit mentions	46.536
Percentage of implicit mentions	53.464
Avg. # of explicit topics per article	1.241
Avg. # of implicit topics per article	1.451
Number of articles without a single topic	24

To compute the F1 score, we first sum the respective TP, FP, and FN values across all topics and then plug them into the F1 equation to get our micro F1 score.

6.2 Auxiliary Information Generation by Simulating Real Users

We observed that in cases in which topic names are not directly mentioned in the article text, one or more informative keywords related to the topic are always present. Indeed, each topic can be conceptually viewed as a word cloud of its informative keywords and different topics will essentially yield different word clouds. In zero-shot learning, these informative keywords (word cloud) are provided by the end user (domain experts) conducting the categorization task. We realized this is what happens often in real-world use cases and decided to simulate this scenario artificially. We extracted potential informative keywords/phrases for each topic using the TF-IDF (term frequency—Inverse Document Frequency) heuristics. From these items, we selected some appropriate words/phrases through manual inspection in order to use them as auxiliary information for the corresponding topic. For example, the articles related to the topic “Gene” yielded informative keywords such as “IL-8,”“TNF-alpha, ”“miR-93, ”and“IFN-gamma.” This way, we prepared a lookup dictionary with individual topics and their respective auxiliary information. We would like to highlight that, using TF-IDF, we first extracted the top 30 keywords for each topic. Next, to simulate the practical scenario, we selected a maximum of 6 to 7 keywords among them for each topic. What follow are a few reasons why we contemplated limiting the keyword count.

- The user conducting the categorization may be unable to provide a comprehensive list of keywords.
- It is impossible to provide an exhaustive list of all related keywords for each topic.
- The idea of keywords is introduced to help/guide in the zero-shot categorization process, not to be exclusively limited to the auxiliary information.

Therefore, we can say that the performance of the categorization will not suffer severely even if the user feeds a small subset of keywords. The topics and the respective keywords are given in Table 4.

6.3 Performance Analysis of Zero-shot Methods

This section presents the results achieved by different zero-shot categorization methods. Table 5 summarizes different topics that are modeling based and classical embedding-based zero-shot methods as well as their hybrids. We noticed that GFLM-Sentence-Moderate and GFLM-Word-Moderate performed better than GFLM-Word and GFLM-Sentence baseline algorithms in terms of Precision, Recall, and F_1-Measure. In fact, among the topic modeling-based approach, GFLM-Word-Moderate achieved the best F_1 score of 0.477. In the case of classical embeddings-based methods, Cosine-Word performed better than other embedding-based methods, achieving F_1 Measure of

Table 4. Topics and Respective Auxiliary Information

Topic	Keywords
CellLine	Huh-7, RaTG13, A549, E6, VeroE6, Caco-2, HeLa
Chemical	Oxygen, Hydroxychloroquine, Remdesivir, Lopinavir, Ritonavir, Oseltamivir
Disease	Anxiety, Cough, Coronavirus, Fever, Fatigue, Pneumonia
Gene	IL-8, TNF-alpha, miR-93, IFN-gamma, TMPRSS2, IL-1beta, RdRp
Mutation	D614G, D936Y, G20210A, L84S, V483A, N501T
Species	MERS-CoV, 2019-nCoV, HCoV-OC43, SARS-CoV, SARS-CoV-2

Table 5. Performance Comparison of Different Topic-Based Approaches such as GFLM-Sentence (GFLM-S), GFLM-Word (GFLM-W), GFLM-Sentence-Moderate (GFLM-S-M), and GFLM-Word-Moderate (GFLM-W-M)

Ad-Hoc Topic Inference											
GFLM-S			**GFLM-W**			**GFLM-S-M**			**GFLM-W-M**		
Precision	**Recall**	F_1	**Precision**	**Recall**	F_1	**Precision**	**Recall**	F_1	**Precision**	**Recall**	F_1
0.397	0.290	0.335	0.393	0.337	0.363	0.511	0.363	0.425	0.526	0.436	**0.477**
Classical Embedding Approaches											
Euclidean-Sentence			**Euclidean-Word**			**Cosine-Sentence**			**Cosine-Word**		
Precision	**Recall**	F_1	**Precision**	**Recall**	F_1	**Precision**	**Recall**	F_1	**Precision**	**Recall**	F_1
0.532	0.993	0.693	0.451	0.998	0.621	0.891	0.571	0.696	0.637	0.774	**0.697**
Hybrid Approaches											
Hybrid S (Euclidean)			**Hybrid W (Euclidean)**			**Hybrid S (Cosine)**			**Hybrid W (Cosine)**		
Precision	**Recall**	F_1	**Precision**	**Recall**	F_1	**Precision**	**Recall**	F_1	**Precision**	**Recall**	F_1
0.408	0.798	0.540	0.462	0.267	0.338	0.411	0.795	**0.542**	0.462	0.267	0.285

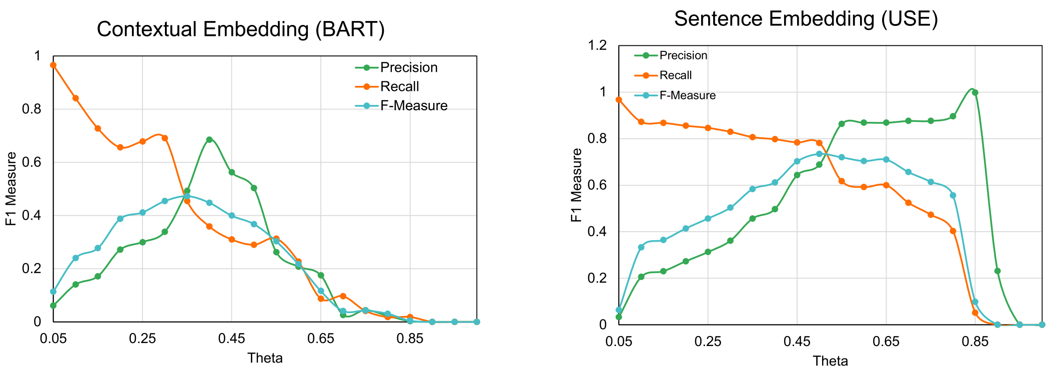
Euclidean-Sentence, Euclidean-Word, Cosine-Sentence, and Cosine-Word denote classical embedding-based approaches in which distance measure is used as Euclidean or cosine measure. Hybrid S (Euclidean), Hybrid W (Euclidean), Hybrid S (cosine), and, Hybrid W (cosine) represent Joint Topic and Embedding-Based Zero-Shot methods in which distance measure is used as Euclidean or cosine measure.

Table 6. Performance Comparison of Different Context Embeddings-Based Approaches such as BERT, ELMO, BART, and Sentence Embedding-Based Approaches such as Universal Sentence Encoder (USE), Sentence-BERT (SBERT), InferSent Encoder

Contextual Embedding Approaches								
BERT			**ELMO**			**BART**		
Precision	**Recall**	F_1	**Precision**	**Recall**	F_1	**Precision**	**Recall**	F_1
0.448	0.999	0.619	0.558	0.987	**0.713**	0.493	0.455	0.473
Sentence Embedding Approaches								
USE			**SBERT**			**InferSent**		
Precision	**Recall**	F_1	**Precision**	**Recall**	F_1	**Precision**	**Recall**	F_1
0.864	0.721	**0.786**	0.451	0.982	0.618	0.229	0.302	0.447

0.697. When it comes to hybrid approaches, in which we merged classical embeddings and topic-based approaches, we observed that Hybrid Sentence Cosine performed better than basic topic inference methods, obtaining an F_1 Measure of 0.54 (corresponding Recall of 0.79 and Precision of 0.411). However, it could not outperform classical embedding-based approaches, which was indeed surprising!

In Table 6, we summarized the performances of more recent embedding techniques such as *Contextual Word Embeddings* (BERT, ELMO, BART) and *Sentence Embeddings* (USE, SBERT, InferSent) based zero-shot approaches. In comparison with BERT and BART, the ELMO-based categorization



(a) Performance analysis of BART

(b) Performance analysis of USE

Fig. 6. Threshold sensitivity (performance over various thresholds).

method obtained better results (F_1 Measure of 0.713), whereas BERT and BART achieved F_1 Measures of 0.619 and 0.473 only. Among Sentence embedding techniques, USE outperformed all other methods with a decent Precision (0.846) and Recall (0.721) score. InferSent performed poorly (F_1 score of 0.447), whereas SBERT was mediocre (0.618). Two sample threshold sensitivity graphs are presented in Figure 6(a) and 6(b) (other plots are omitted due to lack of space).

Due to the very large size of the dataset, this evaluation on zero-shot categorization was performed on the subset of the whole dataset containing 7,000 random samples. Based on the performance results reported in Tables 5 and 6, we picked USEs as our ultimate zero-shot categorization technique due to their highest performance and used their labeled topics for subsequent spatio-temporal analysis tasks.

7 SPATIO-TEMPORAL VISUALIZATION

Once we categorized all research articles, the next step was to use them for spatio-temporal analysis. Spatio-temporal analyses allow the investigator to simultaneously study the persistence of patterns over different times and spaces (i.e., locations), and illuminate interesting patterns. Considering the huge volume of research articles around the world, spatio-temporal analysis is particularly helpful for analyzing trends of research. The COVID Research Tracker uses Google maps, place API, reverse geo-coding, place picker, map cluster, marker pointer, and a few other key features. In this section, we will discuss the different spatio-temporal visualizations of COVID-19 research trends that we created as part of the tracker.

7.1 Spatial Visualization

7.1.1 Clustered Research-Count Map. Figure 7 shows our spatial visualization—clustered research map. *Folium* [29], along with the *markercluster* plugin, was used to produce this map. It shows the frequency distribution of COVID-19 research topics around the world. Research articles mentioning one of the six different topics and researchers' locations are included in this map. This map also provides dynamic clustered representations of research topics on different zoom levels. The clusters are divided into multiple sub-clusters or merged into a bigger cluster upon zooming in or out. This functionality enables users to observe the trend of research in various locations with variable granularity.

The number on each marker cluster icon denotes the number of associated research articles in that cluster. The user can get detailed information about a cluster by interactively hovering

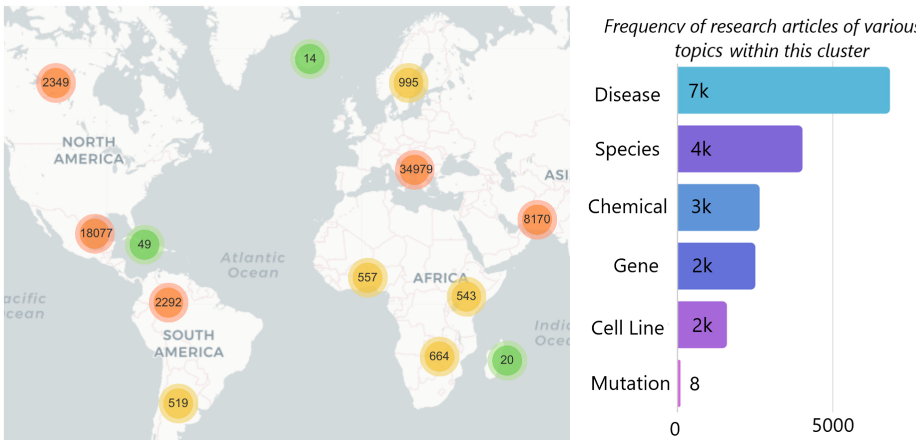
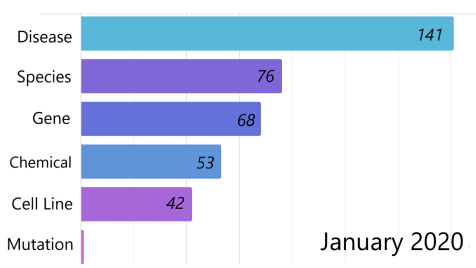
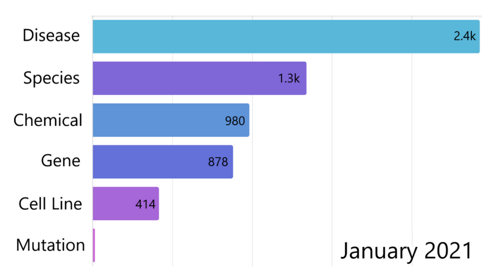


Fig. 7. Clustered research article count over countries



(a) Dynamic Temporal Simulation: Worldwide research trend in January 2020



(b) Dynamic Temporal Simulation: Worldwide research trend in January 2021

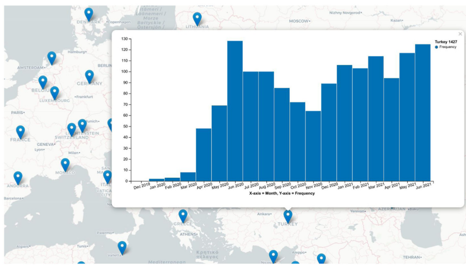
Fig. 8. Temporal simulation.

over the cluster icon, as shown in Figure 7. This action prompts a pop-up window providing a comprehensive summary of that cluster, i.e., the frequency for each topic within that cluster. Nearby points are displayed in a single spiderfy instead of multiple markers to avoid the problem of overlapping markers. Using this visualization, one can easily observe the ongoing COVID-19 research trends in any part of the world—especially through real-time visualization. For example, at the continent level, we found that Europe is much more active in terms of publishing research papers than the North American continent. On the other hand, on the country level, the United States outperformed others in terms of the number of published research papers.

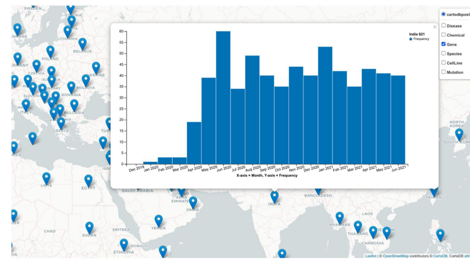
7.2 Temporal Visualization

We created two different kinds of dynamic temporal simulation, which are discussed in this section. Note that, for temporal visualization, we fix the location (e.g., the whole world or a particular country) and vary the time in order to capture the temporal trend.

7.2.1 Dynamic Temporal Simulation. The first temporal visualization uses the monthly aggregated data as shown in Figure 8(a) and 8(b). Using this visualization, users are able to understand worldwide popular research trends over time. Here, we observed a few interesting movements: (a) "Disease," "Species,"and "Chemical" are the most discussed categories over the period of the



(a) Time-series Map: Research article count over time for a specific country



(b) Time-series Map: Research article count for a country (for specific topic)

Fig. 9. Time-series Map: Research article count over time for a specific country.

collected dataset; (b) the Topic “Disease” is dominant over all other research areas in most of the countries, and (c) research in other categories such as “Chemical”became more prominent after April 2020. Users can also observe the dynamic changes in research trends through frequencies of research topics over time.

7.2.2 Time-series Map. The second type of temporal visualization shows country-level time-series analysis. We intend to capture the drift of COVID-19 research in a particular country through this visualization (refer to Figures 9(a) and 9(b)). Here, we show two similar time-series visualization maps with slightly different types of data. In both figures, each location marker on the maps denotes a single country, island, or sea location representing information about country-level aggregated articles. First, Figure 9(a) shows a time-series map for all COVID-related research articles demonstrating how the total number of COVID-19 articles varied over time. One can hover over the marker to find the total number of articles posted about COVID-19 in that specific location. Upon clicking on a marker, a pop-up appears depicting the time series (bar plot) of the COVID-19 research over the specified period. Second, Figure 9(b) shows a more customized time-series map, in which a user can select a specific topic(s) using a checkbox menu to create a custom time series of the user’s choice. It renders a clear idea about the trend or dominance of a particular set of topics over a time span for a specific country. In the time-series plot, the X-axis denotes the date and the Y-axis represents the number of articles related to the particular topic for that period.[4]

From these time-series maps, we find that there was a sudden uprise in the number of COVID-19-related articles during mid-2020, which persisted for nearly a year.

7.3 Visualizing Evolution of Topics Across Time and Space Jointly

So far, we have discussed visualizations to observe the evolution of topics over one of the two dimensions—time or space. We now focus on visualizing the evolution of research topics over time and space jointly. We created two types of visualizations for this purpose. First, given two topics of the user’s choice, the system will generate a juxtapositioned view of how these topic-related article counts changed over time and across geographic locations simultaneously through dynamic bar charts (see Figure 10).

First, we have seen overal that, during the initial phase of the COVID-19 outbreak, China dominated the research field in publishing COVID-19-related articles more than any other country for

[4]We used Python *Folium* [29] to plot the coordinates on the map and the *Vincent* (https://vincent.readthedocs.io/) plugin to produce a time-series plot for each geographic location.

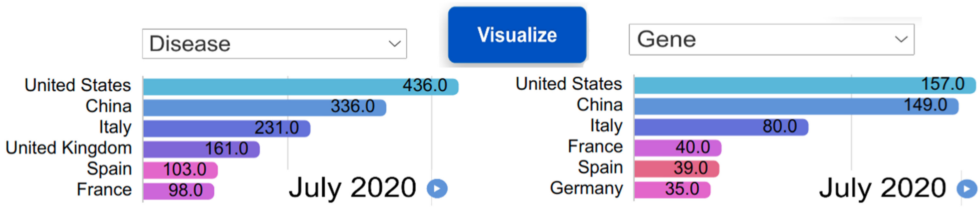


Fig. 10. Time-series map: Spatio-temporal evolution between two topics.

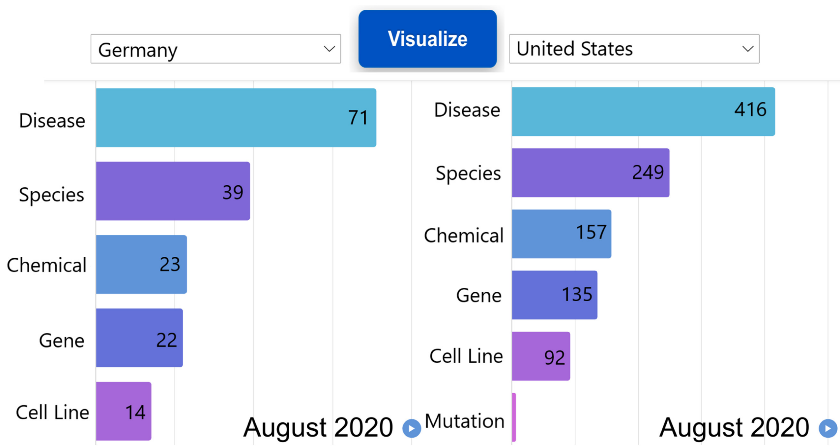


Fig. 11. Time-series map: Temporal evolution of research topics between two countries.

any topic. This is expected as COVID-19 was discovered first in China. However, in the later period (mid-2020), we see that the United States emerged as a pioneer in uncovering unknown factors and preventive measures. Second, a similar visualization, such as that in Figure 10, now with two geographic locations selected by the user from a drop-down menu, shows dynamic changes in topics over time between the two selected locations (see Figure 11).

8 LIMITATIONS

In this section, we will discuss some limitations of our work.

(1) Presently, we do not have any domain experts, policy makers, and funding agencies involved in this study; we will work in the future to mitigate this.
(2) The topics in this article might seem broad, whereas a user may be interested in more detailed categorization. While a more detailed categorization is certainly possible, the reason behind focusing on broad topics in our experiments is as follows:
   - As a major contribution to this article, we have implemented and evaluated several Zero-Shot Topic Categorization methods (topic-based, embedding-based, and transformer-based) and have shown performance comparison of different zero-shot models in Tables 4 and 5. Evaluation/performance comparison of different models is not possible without ground-truth labels. Therefore, to rigorously evaluate our zero-shot topic categorization models, we needed an annotated dataset that contains the ground-truth labels for a large number of text articles. Since CORD-19 and PubTator Central are both well-cited standard datasets for COVID-19, we decided to leverage them. We observed that the

PubTator Central dataset contains full-text articles with the categorization labels of these six biomedical topics; hence, we decided to categorize all of the articles with these six topics.

- However, the Ad-Hoc Topic Tracking system can work with any topics provided by policy makers. The reason is that our proposed zero-shot models are fairly general; upon giving a set of documents, a set of topics, and some auxiliary information (optional), it can categorize any documents. Therefore, policy makers are free to choose any topics that will help them in analyzing the COVID-19 research trend.

It is noteworthy that these limitations do not hurt the general applicability of the proposed technique. That being said, *Ad-Hoc Topic Tracking* can perform up to snuff on entirely different datasets and topics of interest.

9 DISCUSSION AND CONCLUSION

In this article, we built an ad-hoc topic-tracking system to interactively visualize the spatio-temporal evolution of COVID-19-related research by analyzing a huge corpus of research literature. This tool can empower policy makers and funding agencies to better understand such trends using different spatio-temporal visualizations on an ad-hoc basis. We believe this tool will contribute to better-targeted policies by enabling a well-informed decision-making process.

As the ultimate goal of any intelligent tool is to serve the needs of the end users, it is very important to focus on the real-world application scenarios involving the end users. As such, this article contributes towards an ad-hoc concept tracking approach that is mostly unsupervised in nature and can serve end users such as policy makers and sponsors. To accomplish this, we first performed an exhaustive study of zero-shot topic categorization methods, which can be utilized in an ad-hoc fashion for categorizing topics. Most importantly, the benefit of using a zero-shot topic categorization method is that the users can now define their own topics of interest and are not bound by any predefined set of topics. This will immensely help experts to define topic-related custom metadata for their own datasets. We evaluated the zero-shot topic categorization methods on a gold standard CORD-19/PubTator Central dataset, and the zero-shot categorization methods developed in this work are very general. Therefore, it can also be applied to any kind of text data for similar purposes.

In addition, this study provides an overview of the spatio-temporal patterns of COVID-19 research around the globe based on the labeled datasets. Spatial analysis indicated a pattern of spatial clustering of COVID-19 research trends across different countries whereas the temporal analysis presents research transition over time. Considering both time and space, we presented research trends over time and different geographic locations. The tool is available online for public use at https://bijoy-sust.github.io/Annotation/index.html.

REFERENCES

[1] Tauseef Ahmad, Manal Abdulaziz Murad, Mukhtiar Baig, and Jin Hui. 2021. Research trends in COVID-19 vaccine: A bibliometric analysis. *Human Vaccines & Immunotherapeutics* 17, 8 (2021), 2367–2372.

[2] Wasim Ahmed, Josep Vidal-Alaball, Joseph Downing, and Francesc López Seguí. 2020. COVID-19 and the 5G conspiracy theory: Social network analysis of Twitter data. *Journal of Medical Internet Research* 22, 5 (2020), e19458.

[3] Firoj Alam, Shaden Shaar, Fahim Dalvi, Hassan Sajjad, Alex Nikolov, Hamdy Mubarak, Giovanni Da San Martino, Ahmed Abdelali, Nadir Durrani, Kareem Darwish, Abdulaziz Al-Homaid, Wajdi Zaghouani, Tommaso Caselli, Gijs Danoe, Friso Stolk, Britt Bruntink, and Preslav Nakov. 2021. Fighting the COVID-19 infodemic: Modeling the perspective of journalists, fact-checkers, social media platforms, policy makers, and the society. In *Findings of the Association for Computational Linguistics: EMNLP 2021, Virtual Event / Punta Cana, Dominican Republic, 16–20 November, 2021*, Marie-Francine Moens, Xuanjing Huang, Lucia Specia, and Scott Wen-tau Yih (Eds.). Association for Computational Linguistics, 611–649. https://doi.org/10.18653/v1/2021.findings-emnlp.56

[4] Daniel M. Altmann, Daniel C. Douek, and Rosemary J. Boyton. 2020. What policy makers need to know about COVID-19 protective immunity. *The Lancet* 395, 10236 (2020), 1527–1529.

[5] Abhinav Arora, Akshat Shrivastava, Mrinal Mohit, Lorena Sainz-Maza Lecanda, and Ahmed Aly. 2020. Cross-lingual transfer learning for intent detection of COVID-19 utterances. (2020).

[6] Rachel Atherton. 2021. "Missing/Unspecified": Demographic data visualization during the COVID-19 pandemic. *Journal of Business and Technical Communication* 35, 1 (2021), 80–87.

[7] Thirunavukarasu Balasubramaniam, Richi Nayak, and Md. Abul Bashar. 2020. Understanding the spatio-temporal topic dynamics of COVID-19 using nonnegative tensor factorization: A case study. *CoRR* abs/2009.09253 (2020). arXiv:2009.09253 https://arxiv.org/abs/2009.09253.

[8] Sunil Bhopal, Jay Bagaria, and Raj Bhopal. 2020. Children's mortality from COVID-19 compared with all-deaths and other relevant causes of death: Epidemiological information for decision-making by parents, teachers, clinicians and policymakers. *Public Health* 185 (2020), 19.

[9] Biddut Sarker Bijoy, Syeda Jannatus Saba, Souvika Sarkar, Md Saiful Islam, Sheikh Rabiul Islam, Mohammad Ruhul Amin, and Shubhra Kanti Karmaker Santu. 2021. COVID19α: Interactive spatio-temporal visualization of COVID-19 symptoms through tweet analysis. In *IUI'21: 26th International Conference on Intelligent User Interfaces, College Station, TX, USA, April 13–17, 2021, Companion*, Tracy Hammond, Katrien Verbert, and Dennis Parra (Eds.). ACM, 28–30. https://doi.org/10.1145/3397482.3450715

[10] David M. Blei, Andrew Y. Ng, and Michael I. Jordan. 2003. Latent Dirichlet allocation. *The Journal of Machine Learning Research* 3 (2003), 993–1022.

[11] Piotr Bojanowski, Edouard Grave, Armand Joulin, and Tomas Mikolov. 2017. Enriching word vectors with subword information. *Transactions of the Association for Computational Linguistics* 5 (2017), 135–146.

[12] Kalina Bontcheva, Hamish Cunningham, Ian Roberts, Angus Roberts, Valentin Tablan, Niraj Aswani, and Genevieve Gorrell. 2013. GATE teamware: A web-based, collaborative text annotation framework. *Language Resources and Evaluation* 47, 4 (2013), 1007–1029.

[13] J. Scott Brennen, Felix Simon, Philip N. Howard, and Rasmus Kleis Nielsen. 2020. Types, sources, and claims of COVID-19 misinformation. *Reuters Institute* 7 (2020), 3–1.

[14] Álvaro Briz-Redón and Ángel Serrano-Aroca. 2020. A spatio-temporal analysis for exploring the effect of temperature on COVID-19 early evolution in Spain. *Science of the Total Environment* 728 (2020), 138811.

[15] Henna Budhwani and Ruoyan Sun. 2020. Creating COVID-19 stigma by referencing the novel coronavirus as the "Chinese virus" on Twitter: Quantitative analysis of social media data. *Journal of Medical Internet Research* 22, 5 (2020), e19301.

[16] Markus Bundschus, Volker Tresp, and Hans-Peter Kriegel. 2009. Topic models for semantically annotated document collections. In *NIPS Workshop: Applications for Topic Models: Text and Beyond*. 1–4.

[17] Daniel Cer, Yinfei Yang, Sheng-yi Kong, Nan Hua, Nicole Limtiaco, Rhomni St John, Noah Constant, Mario Guajardo-Cespedes, Steve Yuan, Chris Tar, Brian Strope, and Ray Kurzweil. 2018. Universal sentence encoder for English. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, Association for Computational Linguistics, Brussels, 169–174. https://aclanthology.org/D18-2029.

[18] Ilias Chalkidis, Manos Fergadiotis, Sotiris Kotitsas, Prodromos Malakasiotis, Nikolaos Aletras, and Ion Androutsopoulos. 2020. An empirical study on large-scale multi-label text classification including few and zero-shot labels. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP'20), Online, November 16-20, 2020*, Bonnie Webber, Trevor Cohn, Yulan He, and Yang Liu (Eds.). Association for Computational Linguistics, 7503–7515. https://doi.org/10.18653/v1/2020.emnlp-main.607

[19] Shakti Chaturvedi and Thomas Enias Pasipanodya. 2021. A perspective on reprioritizing children's' wellbeing amidst COVID-19: Implications for policymakers and caregivers. *Frontiers in Human Dynamics* 2 (2021), 18.

[20] Baoquan Chen, Mingyi Shi, Xingyu Ni, Liangwang Ruan, Hongda Jiang, Heyuan Yao, Mengdi Wang, Zhenghua Song, Qiang Zhou, and Tong Ge. 2020. Data visualization analysis and simulation prediction for COVID-19. *arXiv preprint arXiv:2002.07096* (2020).

[21] Qi Chen, Wei Wang, Kaizhu Huang, and Frans Coenen. 2021. Zero-shot text classification via knowledge graph embedding for social media data. *IEEE Internet of Things Journal* (2021).

[22] Alexis Conneau, Douwe Kiela, Holger Schwenk, Loic Barrault, and Antoine Bordes. 2017. Supervised learning of universal sentence representations from natural language inference data. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing (EMNLP'17, Copenhagen, Denmark, September 9-11, 2017)*, Association for Computational Linguistics, 670–680. https://doi.org/10.18653/v1/d17-1070

[23] Steven J. Cooke, Peter Soroye, Jill L. Brooks, Julia Clarke, Amanda L. Jeanson, Albana Berberi, Morgan L. Piczak, Connor H. Reid, Jessica E. Desforges, J. D. Guay, et al. 2021. Ten considerations for conservation policy makers for the post-COVID-19 transition. *Environmental Reviews* 29, 999 (2021), 1–8.

[24] Stephen Dill, Nadav Eiron, David Gibson, Daniel Gruhl, Ramanathan Guha, Anant Jhingran, Tapas Kanungo, Sridhar Rajagopalan, Andrew Tomkins, John A. Tomlin, et al. 2003. SemTag and Seeker: Bootstrapping the semantic web via automated semantic annotation. In *Proceedings of the 12th International Conference on World Wide Web*. 178–186.

[25] Ram A. Dixit, Stephen Hurst, Katharine T. Adams, Christian Boxley, Kristi Lysen-Hendershot, Sonita S. Bennett, Ethan Booker, and Raj M. Ratwani. 2020. Rapid development of visualization dashboards to enhance situation awareness of COVID-19 telehealth initiatives at a multihospital healthcare system. *Journal of the American Medical Informatics Association* 27, 9 (2020), 1456–1461.

[26] Lan Du, Wray Buntine, and Mark Johnson. 2013. Topic segmentation with a structured topic model. In *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. ACL, 190–200.

[27] Chris Engels, Koen Deschacht, Jan Hendrik Becker, Tinne Tuytelaars, Sien Moens, and Luc J. Van Gool. 2010. Automatic annotation of unique locations from video and text. In *BMVC*. 1–11.

[28] Xi Fang, Uwe Kruger, Fatemeh Homayounieh, Hanqing Chao, Jiajin Zhang, Subba R. Digumarthy, Chiara D. Arru, Mannudeep K. Kalra, and Pingkun Yan. 2021. Association of AI quantified COVID-19 chest CT and patient outcome. *International Journal of Computer Assisted Radiology and Surgery* 16, 3 (2021), 435–445.

[29] Filipe, Martin Journois, Frank, Rob Story, James Gardiner, Halfdan Rump, Andrew Bird, Antonio Lima, Joshua Cano, Juliana Leonel, Tim Sampson, Ben Welsh, Jon Reades, Oleg Komarov, Jason Baker, Qingkai Kong, odovad, Raphael Dumas, George Harris, Alex Crosby, kenmatsu4, Tales Paiva Nogueira, Nat Wilson, Daisuke Kato, andrew giessel, soymsk, Rich Signell, Justin Duke, Anand Patil, and FabeG. 2019. *python-visualization/folium: v0.9.1*. Retrieved March 10, 2020 from https://doi.org/10.5281/zenodo.3229045

[30] Peter Fotheringham, Thomas Harriott, Grace Healy, Gabrielle Arenge, Ross McGill, and Elaine Wilson. 2020. Pressures and influences on school leaders as policy makers during COVID-19. *Available at SSRN 3642919* (2020).

[31] Nicholas Fraser, Liam Brierley, Gautam Dey, Jessica K. Polka, Máté Pálfy, Federico Nanni, and Jonathon Alexis Coates. 2021. The evolving role of preprints in the dissemination of COVID-19 research and their impact on the science communication landscape. *PLOS Biology* 19, 4 (04 2021), 1–28. https://doi.org/10.1371/journal.pbio.3000959

[32] Mahaveer Golechha. 2020. COVID-19 containment in Asia's largest urban slum Dharavi-Mumbai, India: Lessons for policymakers globally. *Journal of Urban Health* 97, 6 (2020), 796–801.

[33] Meenu Gupta, Rachna Jain, Simrann Arora, Akash Gupta, Mazhar Javed Awan, Gopal Chaudhary, and Haitham Nobanee. 2021. AI-enabled COVID-19 outbreak analysis and prediction: Indian states vs. union territories. *Computers, Materials and Continua* 67, 1 (2021).

[34] F. A. Binti Hamzah, C. Lau, H. Nazri, D. V. Ligot, G. Lee, C. L. Tan, M. K. B. M. Shaib, U. H. B. Zaidon, A. B. Abdullah, M. H. Chung, et al. 2020. CoronaTracker: Worldwide COVID-19 outbreak data analysis and prediction. *Bull World Health Organ* 1 (2020), 32.

[35] Mainul Haque. 2020. Combating COVID-19: A coordinated efforts of healthcare providers and policy makers with global participation are needed to achieve the desired goals. *Bangladesh Journal of Medical Science* (2020), 01–05.

[36] L. Harper, N. Kalfa, G. M. A. Beckers, M. Kaefer, A. J. Nieuwhof-Leppink, Magdalena Fossum, K. W. Herbst, D. Bagli, ESPU Research Committee, et al. 2020. The impact of COVID-19 on research. *Journal of Pediatric Urology* 16, 5 (2020), 715.

[37] Tieke He, Hongzhi Yin, Zhenyu Chen, Xiaofang Zhou, Shazia Sadiq, and Bin Luo. 2016. A spatial-temporal topic model for the semantic annotation of POIs in LBSNs. *ACM Transactions on Intelligent Systems and Technology* 8, 1 (2016), 1–24.

[38] Adedoyin Ahmed Hussain, Ouns Bouachir, Fadi Al-Turjman, and Moayad Aloqaily. 2020. AI techniques for COVID-19. *IEEE Access* 8 (2020), 128776–128795.

[39] Hans IJzerman, Neil A. Lewis, Andrew K. Przybylski, Netta Weinstein, Lisa DeBruine, Stuart J. Ritchie, Simine Vazire, Patrick S. Forscher, Richard D. Morey, James D. Ivory, et al. 2020. Use caution when applying behavioural science to policy. *Nature Human Behaviour* 4, 11 (2020), 1092–1094.

[40] Tomoharu Iwata, Takeshi Yamada, and Naonori Ueda. 2009. Modeling social annotation data with content relevance using a topic model. In *Advances in Neural Information Processing Systems*. 835–843.

[41] Debanjana Kar, Mohit Bhardwaj, Suranjana Samanta, and Amar Prakash Azad. 2020. No rumours please! A multi-indic-lingual approach for COVID fake-tweet detection. In *2021 Grace Hopper Celebration India (GHCI)*. IEEE, 1–5.

[42] Shubhra Kanti Karmaker Santu, Parikshit Sondhi, and ChengXiang Zhai. 2016. Generative feature language models for mining implicit features from customer reviews. In *Proceedings of the 25th ACM International Conference on Information and Knowledge Management*. 929–938.

[43] Ramez Kouzy, Joseph Abi Jaoude, Afif Kraitem, Molly B. El Alam, Basil Karam, Elio Adib, Jabra Zarka, Cindy Traboulsi, Elie W. Akl, and Khalil Baddour. 2020. Coronavirus goes viral: Quantifying the COVID-19 misinformation epidemic on Twitter. *Cureus* 12, 3 (2020).

[44] Yanis Labrak and Richard Dufour. 2021. Team LIA/LS2N at BioCreative VII LitCovid track: Multi-label document classification for COVID-19 literature using keyword based enhancement and few-shot learning. In *BioCreative VII Challenge Evaluation Workshop*.

[45] Sijia Li, Yilin Wang, Jia Xue, Nan Zhao, and Tingshao Zhu. 2020. The impact of COVID-19 epidemic declaration on psychological consequences: A study on active Weibo users. *International Journal of Environmental Research and Public Health* 17, 6 (2020), 2032.

[46] Yuyu Luo, Wenbo Li, Tianyu Zhao, Xiang Yu, Lixi Zhang, Guoliang Li, and Nan Tang. 2020. DeepTrack: Monitoring and exploring spatio-temporal data: A case of tracking COVID-19. *Proceedings of the VLDB Endowment* 13, 12 (2020), 2841–2844.

[47] Simon Lupart, Benoit Favre, Vassilina Nikoulina, and Salah Ait-Mokhtar. 2022. Zero-shot and few-shot classification of biomedical articles in context of the COVID-19 pandemic. *arXiv preprint arXiv:2201.03017* (2022).

[48] Sean MacAvaney, Arman Cohan, and Nazli Goharian. 2020. SLEDGE-Z: A zero-shot baseline for COVID-19 literature search. *arXiv preprint arXiv:2010.05987* (2020).

[49] Noveri Maulana. 2020. Research trends in marketing science before COVID-19 outbreak: A literature review. *Management & Marketing* 15 (2020), 514–533.

[50] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781* (2013).

[51] Vicki Moulder, Lorna R. Boschman, Ron Wakkary, Carman Neustaedter, and Hiroki Hill Kobayashi. 2018. HCI interventions for science communication. In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems*. 1–9.

[52] National Academies of Sciences. 2017. Communicating science effectively: A research agenda. (2017).

[53] Philip V. Ogren. 2006. Knowtator: A protégé plug-in for annotated corpus construction. In *Proceeding of the Human Language Technology Conference of the North American Chapter of the Association of Computational Linguistics, Proceedings, June 4-9, 2006, New York, New York, USA)*, Robert C. Moore, Jeff A. Bilmes, Jennifer Chu-Carroll, and Mark Sanderson, The Association for Computational Linguistics. https://aclanthology.org/N06-4006/.

[54] Antonio Paez, Fernando A. Lopez, Tatiane Menezes, Renata Cavalcanti, and Maira Galdino da Rocha Pitta. 2020. A spatio-temporal analysis of the environmental correlates of COVID-19 incidence in Spain. *Geographical Analysis* (2020).

[55] Jeffrey Pennington, Richard Socher, and Christopher D. Manning. 2014. Glove: Global vectors for word representation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP'14)*. 1532–1543.

[56] Forough Poursabzi-Sangdeh and Jordan Boyd-Graber. 2015. Speeding document annotation with topic models. In *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Student Research Workshop*. 126–132.

[57] Raul Puri and Bryan Catanzaro. 2019. Zero-shot text classification with generative language models. *arXiv preprint arXiv:1912.10165* (2019).

[58] Pushpankar Kumar Pushp and Muktabh Mayank Srivastava. 2017. Train once, test anywhere: Zero-shot learning for text classification. *arXiv preprint arXiv:1712.05972* (2017).

[59] Nils Reimers and Iryna Gurevych. 2019. Sentence-BERT: Sentence embeddings using Siamese BERT-networks. *arXiv preprint arXiv:1908.10084* (2019).

[60] Mahdi Rezaei and Mahsa Shahidi. 2020. Zero-shot learning and its applications from autonomous vehicles to COVID-19 diagnosis: A review. *Intelligence-based Medicine* 3 (2020), 100005.

[61] Anthony Rios and Ramakanth Kavuluru. 2018. Few-shot and zero-shot multi-label learning for structured label spaces. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, Vol. 2018. NIH Public Access, 3132.

[62] Anit N. Roy, Jais Jose, Aswin Sunil, Neha Gautam, Deepa Nathalia, and Arjun Suresh. 2020. Prediction and spread visualization of Covid-19 pandemic using machine learning.

[63] Souvika Sarkar and Shubhra Kanti Karmaker Santu. 2022. Concept annotation from users perspective: A new challenge. In *Companion of The Web Conference 2022, Virtual Event / Lyon, France, April 25 - 29, 2022*, Frédérique Laforest, Raphaël Troncy, Elena Simperl, Deepak Agarwal, Aristides Gionis, Ivan Herman, and Lionel Médini (Eds.). ACM, 1180–1188. https://doi.org/10.1145/3487553.3524933

[64] Huang Sen-zhong, Peng Zhihang, and Jin Zhen. 2020. Studies of the strategies for controlling the COVID-19 epidemic in China: Estimation of control efficacy and suggestions for policy makers. *Scientia Sinica Mathematica* 50, 6 (2020), 885.

[65] Dexuan Sha, Xin Miao, Hai Lan, Kathleen Stewart, Shiyang Ruan, Yifei Tian, Yuyang Tian, and Chaowei Yang. 2020. Spatiotemporal analysis of medical resource deficiencies in the US under COVID-19 pandemic. *PloS One* 15, 10 (2020), e0240348.

[66] Mohsen Shariati, Tahoora Mesgari, Mahboobeh Kasraee, and Mahsa Jahangiri-Rad. 2020. Spatiotemporal analysis and hotspots detection of COVID-19 using geographic information system (March and April, 2020). *Journal of Environmental Health Science and Engineering* 18, 2 (2020), 1499–1507.

[67] Dr Sharma. 2020. COVID-19 (an international trauma): A brief analysis on research trends, impacts and solutions. *International Journal For Research in Applied Sciences and Biotechnology* 2 (2020). DOI: 10.31033/ijrasb.7.2.1

[68] Karishma Sharma, Sungyong Seo, Chuizheng Meng, Sirisha Rambhatla, and Yan Liu. 2020. COVID-19 on social media: Analyzing misinformation in Twitter conversations. *arXiv preprint arXiv:2003.12309* (2020).

[69] Pranjali Singh and Amritpal Singh. 2021. Unmasking the masked face using zero-shot learning. In *International Conference on Advanced Network Technologies and Intelligent Computing*. Springer, 563–585.

[70] Janice C. Sipior. 2020. Considerations for development and use of AI in response to COVID-19. *International Journal of Information Management* 55 (2020), 102170.

[71] Sonish Sivarajkumar and Yanshan Wang. 2022. HealthPrompt: A zero-shot learning paradigm for clinical natural language processing. *arXiv preprint arXiv:2203.05061* (2022).

[72] Yangqiu Song, Shyam Upadhyay, Haoruo Peng, Stephen Mayhew, and Dan Roth. 2019. Toward any-language zero-shot topic classification of textual documents. *Artificial Intelligence* 274 (2019), 133–150.

[73] Shashank Srivastava, Igor Labutov, and Tom Mitchell. 2018. Zero-shot learning of classifiers from natural language quantification. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. 306–316.

[74] Lin Tian, Xiuzhen Zhang, and Jey Han Lau. 2021. Rumour detection via zero-shot cross-lingual transfer learning. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*. Springer, 603–618.

[75] Suppawong Tuarob, Line C. Pouchard, Prasenjit Mitra, and C. Lee Giles. 2015. A generalized topic modeling approach for automatic document annotation. *International Journal on Digital Libraries* 16, 2 (2015), 111–128.

[76] Zeliha Kocak Tufan and Bircan Kayaaslan. 2020. Crushing the curve, the role of national and international institutions and policy makers in COVID-19 pandemic. *Turkish Journal of Medical Sciences* 50, SI-1 (2020), 495–508.

[77] Raju Vaishya, Mohd Javaid, Ibrahim Haleem Khan, and Abid Haleem. 2020. Artificial intelligence (AI) applications for COVID-19 pandemic. *Diabetes & Metabolic Syndrome: Clinical Research & Reviews* 14, 4 (2020), 337–339.

[78] Sappadla Prateek Veeranna, Jinseok Nam, E. L. Mencía, and J. Furnkranz. 2016. Using semantic similarity for multi-label zero-shot classification of text documents. In *Proceedings of European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges, Belgium: Elsevier*. 423–428.

[79] Surabhi Verma and Anders Gustafsson. 2020. Investigating the emerging COVID-19 research trends in the field of business and management: A bibliometric analysis approach. *Journal of Business Research* 118 (2020), 253–261.

[80] Hongning Wang, Duo Zhang, and ChengXiang Zhai. 2011. Structural Topic model for latent topical structure analysis. In *ACL*.

[81] Lucy Lu Wang, Kyle Lo, Yoganand Chandrasekhar, Russell Reas, Jiangjiang Yang, Darrin Eide, Kathryn Funk, Rodney Kinney, Ziyang Liu, William Merrill, et al. 2020. CORD-19: The COVID-19 open research dataset. *ArXiv* (2020).

[82] Chih-Hsuan Wei, Alexis Allot, Robert Leaman, and Zhiyong Lu. 2019. PubTator Central: Automated concept annotation for biomedical full text articles. *Nucleic Acids Research* 47, W1 (2019), W587–W593.

[83] Congying Xia, Chenwei Zhang, Xiaohui Yan, Yi Chang, and Philip S. Yu. 2018. Zero-shot user intent detection via capsule neural networks. *arXiv preprint arXiv:1809.00385* (2018).

[84] Wenpeng Yin, Jamaal Hay, and Dan Roth. 2019. Benchmarking zero-shot text classification: Datasets, evaluation and entailment approach. *arXiv preprint arXiv:1909.00161* (2019).

[85] Yian Yin, Jian Gao, Benjamin F. Jones, and Dashun Wang. 2021. Coevolution of policy and science during the pandemic. *Science* 371, 6525 (2021), 128–130.

[86] Dani Yogatama, Chris Dyer, Wang Ling, and Phil Blunsom. 2017. Generative and discriminative text classification with recurrent neural networks. *arXiv preprint arXiv:1703.01898* (2017).

[87] Hongyue Zhang and Rajib Shaw. 2020. Identifying research trends and gaps in the context of COVID-19. *International Journal of Environmental Research and Public Health* 17, 10 (2020), 3370.

[88] Jingqing Zhang, Piyawat Lertvittayakumjorn, and Yike Guo. 2019. Integrating semantic knowledge to tackle zero-shot text classification. *arXiv preprint arXiv:1903.12626* (2019).

Received 10 April 2022; revised 21 September 2022; accepted 20 October 2022