

Contents lists available at [ScienceDirect](https://www.sciencedirect.com)

Information Processing and Management

journal homepage: www.elsevier.com/locate/ipm

Do you see what I see? Images of the COVID-19 pandemic through the lens of Google

Monica Lestari Paramita ^{a,*}, Kalia Orphanou ^b, Evgenia Christoforou ^c,
Jahna Otterbacher ^{b,c}, Frank Hopfgartner ^a

^a Information School, University of Sheffield, United Kingdom

^b Open University of Cyprus, Cyprus

^c CYENS - Centre of Excellence, Nicosia, Cyprus

ARTICLE INFO

Keywords:

Images
Information diversity
Pandemic information
Proprietary search

ABSTRACT

During times of crisis, information access is crucial. Given the opaque processes behind modern search engines, it is important to understand the extent to which the “picture” of the Covid-19 pandemic accessed by users differs. We explore variations in what users “see” concerning the pandemic through Google image search, using a two-step approach. First, we crowdsource a search task to users in four regions of Europe, asking them to help us create a photo documentary of Covid-19 by providing image search queries. Analysing the queries, we find five common themes describing information needs. Next, we study three sources of variation – users’ information needs, their geo-locations and query languages – and analyse their influences on the similarity of results. We find that users see the pandemic differently depending on where they live, as evidenced by the 46% similarity across results. When users expressed a given query in different languages, there was no overlap for most of the results. Our analysis suggests that localisation plays a major role in the (dis)similarity of results, and provides evidence of the diverse “picture” of the pandemic seen through Google.

1. Introduction

The Covid-19 pandemic is proving to be an event without precedent. Health experts are describing the multitude of ways that people and communities are feeling the impact, ranging from confusion, isolation, and feelings of insecurity (Pfefferbaum & North, 2020), to large-scale problems such as alcohol and drug abuse (Clay & Parker, 2020), and increased levels of anxiety and sleep disturbances (Sher, 2020). The World Economic Forum reported that “Covid-19 has changed what we search for online” during lockdown, citing an increase in queries of diverse topics such as staying healthy, financial security, personal hobbies and cooking.¹ Studies have further shown that search engine queries related to Covid-19 can be used to predict outbreaks in the area of origin (Cousins, Cousins, Harris, & Pasquale, 2020) and, more importantly, that search activity correlates with daily confirmed cases and deaths in some areas (Higgins et al., 2020). These studies indicate that users rely significantly on search engines to access information during the pandemic. Furthermore, search queries surrounding the pandemic concern not only facts about the virus, but also the implications on our lives, which likely differ from country to country.

* Corresponding author.

E-mail addresses: m.paramita@sheffield.ac.uk (M.L. Paramita), kalia.orphanou@ouc.ac.cy (K. Orphanou), e.christoforou@cyens.org.cy (E. Christoforou), jahna.otterbacher@ouc.ac.cy (J. Otterbacher), f.hopfgartner@sheffield.ac.uk (F. Hopfgartner).

¹ <https://www.weforum.org/agenda/2020/05/google-trends-search-online-coronavirus-covid-19/>

<https://doi.org/10.1016/j.ipm.2021.102654>

Received 30 November 2020; Received in revised form 19 May 2021; Accepted 26 May 2021

Available online 11 June 2021

0306-4573/© 2021 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license

(<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Proprietary search engines – and in particular, Google – play a leading role in the public’s access to information, processing approximately 90% of all queries executed worldwide,² with this figure rising to 93% for Europe.³ As arguably one of the most complex and opaque information access systems used by the public, Google holds great power over what users see – and what they do not see (Diakopoulos, 2015). As demonstrated in Hannak et al. (2013) and Kliman-Silver, Hannak, Lazer, Wilson, and Mislove (2015), Google search performs personalisation of results based on users’ locations, which resulted in different results being shown to users based in different locations. Another study has focused specifically on investigating the localisation aspect, specifically with regards to the localness of search results, i.e., to what extent the retrieved results were published by local sites for different queries (Ballatore, Graham, & Sen, 2017). However, the research to date has not measured how much these processes impact search results in the context of the pandemic.

There is a growing body of literature suggesting that users are largely unaware of the opaque algorithmic processes, which directly influence the information they access (Koenig, 2020; Powers, 2017). Given users’ trust in search engines, these different results have great impact, from influencing users in their scientific knowledge (Novin & Meyers, 2017), manipulating their voting decisions (Epstein & Robertson, 2015), etc. Since providing access to information in this situation is so crucial, it is important to consider the extent to which what users are seeing differs from place to place.

Visual information – such as that contained in images – has been shown to be crucial in the context of science communication (including messages concerning Covid-19). Images are often interpreted as being closer to the truth as compared to other forms of communication, because of their physical representation of an event such as the Covid-19 pandemic (Brennen, Simon, Howard, & Nielsen, 2020; Messaris & Abraham, 2001). Thus, the current work focuses on *image search results* and in particular Google, as a key mechanism for the public to find visual information sources about the pandemic.

1.1. Goals of the study

This study investigates the diversity in the picture of the pandemic that users see through Google image search, focusing on three sources of variation: the queries of interest to residents of a given geographical location (i.e., prevalent information needs), the location from which the search queries are executed, and the language of the query. We also analyse to what extent the *localisation* rates (i.e., ratio of local results in the search results) affect the variations in image search results when users search for visual information on the pandemic.

In the first part of the study, we utilised crowdsourcing to generate queries for visual information concerning Covid-19. We invited people in four European countries (Great Britain,⁴ Germany, Spain and Italy) to participate. We asked them to create queries for a simulated search task of collecting images to document the pandemic. Through an analysis of the queries, we develop a taxonomy for categorising them. In the second part of the study, given our corpus of queries, we analyse the retrieved search results by specifying different geo-locations of the users. We further examine the influence of thematic categories, users’ locations, and the role of the query language. Our study addresses three research questions:

- RQ1. What image search queries are of interest to people across regions?
- RQ2. How similar are the results presented to different users?
- RQ3. What aspects influence the similarity of results?

To our knowledge, this is the first attempt to understand how proprietary search mediates the “picture” of the Covid-19 pandemic seen by people across locations. The paper is structured as follows. In Section 2 we describe related work. In Section 3, we describe the methodology used for the collection of the queries and for the analysis of the queries and search results. In Section 4, we present the analysis of the queries collected, while in Section 5, we provide the results of the analysis of Google search results. In Section 6, the key findings of this work are discussed and in Section 7, we conclude this paper.

2. Related work

2.1. Web auditing

Various aspects, such as user behaviour or their location, have also been utilised by search algorithms to recommend relevant information that match users’ interests more (Bennett, Radlinski, White, & Yilmaz, 2011; Monzer, Moeller, Helberger, & Eskens, 2020). Whilst this process has been shown to improve search performance, it has also increased the diversity between results seen by users, e.g., in the form of “filter bubbles” (Haim, Arendt, & Scherr, 2017). Thus, recent work used Web auditing to investigate the influence of personalisation to the results presented by search engines to different users. For instance, Kliman-Silver et al. (2015) collected and analysed Google results for 240 queries over 30 days from 59 different GPS coordinates in the US to examine the influence of users’ geolocations on search results. They found that 18%–34% search results may vary based on users’ geo-locations. However, these differ between different types of queries. In a similar vein, Hannak et al. (2013) studied personalisation on a diverse

² <https://gs.statcounter.com/search-engine-market-share>

³ <https://gs.statcounter.com/search-engine-market-share/all/europe>

⁴ We used Great Britain instead of the United Kingdom for consistencies with the search API, further described in Section 3.4.

set of queries across 200 Google user accounts. They estimated that on average, 11.7% of search results differ, due to the effect of personalisation.

Other studies (Haim et al., 2017; Scherr, Haim, & Arendt, 2019) investigated whether users searching for suicide-related queries in Google would be shown the suicide helpline at the top of the results. Based on their findings suicide helpline was not equally displayed to all users but differ based on user geo-locations. Additionally, another study found significant differences between politically-related search pages shown to users who were logged in and their incognito tabs (Robertson et al., 2018). These results indicate the significant influence of personalisation in the results that users see. Instead of auditing the results presented to different users, Ballatore et al. (2017) examined the Google search results from a localisation aspect using capital cities as queries. They found that results describing cities from well-developed areas were often published locally, while results describing cities in the “Global South” were more likely to be more diverse and were often dominated by information published by other countries in the “Global North”.

Web auditing has also been performed by comparing results between different search engines (Kamoun, Maillé, & Tuffin, 2019; Makhortykh, Urman, & Roberto, 2020). Makhortykh et al. (2020) examined how six different search engines present results related to Covid-19 using the query “coronavirus” and found that different search engines prioritised results from different sources (e.g., government resources, or non-official resources such as Reddit or social media). Furthermore, the authors found that even the same search engine might randomise the top 10 results for different users. Jiang (2014) compared results retrieved by Google and Baidu, and found that the results achieved only 6.8% overlap with very little similarity between rankings.

2.2. Information needs and Covid-19

In previous studies, users’ search behaviours during the Covid-19 pandemic have been analysed from an infodemiology approach. For instance, Canchari, Chávez-Bustamante, and Caira-Chuquineyra (2020) used Trends data from searches related to the coronavirus disease during the period of January to May 2020 and identified that “coronavirus” was the most frequent search term, followed by “fever”, “sore throat”, and “cough”. In addition, they found that specific queries such as “covid spread”, “face masks”, “stay home”, were related to the increased severity of the pandemic during that period. Using a similar approach, other works (Cousins et al., 2020; Higgins et al., 2020; Walker, Hopkins, & Surda, 2020) utilising both Google Trends and Baidu Index, found out that terms relating to shortness of breath, headache, chest pain and loss of smell correlate with rates of confirmed cases and deaths. Another study investigated aspects influencing the use of specific queries (Sousa-Pinto et al., 2020). Whilst they found that the term “coronavirus” was used frequently throughout the pandemic, the use of some queries was found to be influenced by media coverage. E.g., queries such as “ageusia” (loss of taste) and “anosmia” (loss of smell) were only found in the search trends once they have been reported as Covid symptoms by the media.

2.3. Summary

Our information landscape is characterised by mediation from opaque algorithmic systems, and in particular, the Google search engine. Many users are entirely unaware that their information choices are under the influence of algorithms. Meanwhile, users who are aware of algorithmic mediation often express frustration with their lack of control over information filter bubbles. However, with only one search engine serving most of the world’s users, it becomes difficult to understand how a search engine “should” behave, and which results are most appropriate for whom. To this end, much recent work has focused on achieving a better understanding of how Google in particular, customises users’ search results, based on their location or profile.

With this in mind, in the next section, we detail the methodology of our study. As will be explained, we focus on studying the impact of three factors – users’ information needs, their location and the source language of their query – on the image URLs and visual content presented to them by Google. By comparing the results across four European countries, we characterise the above three factors as sources of variation in a user’s view on the Covid-19 pandemic.

3. Methodology

3.1. Overview

As shown in Fig. 1, we take a two-step approach, involving a crowdsourcing study and an analysis of search results. First, we aimed to collect, in a natural manner, queries that Google users would pose to search images relevant to the Covid-19 pandemic. Collecting image search queries through crowdsourcing allows us to exploit the wisdom of the crowd (Brabham, 2008; Ranard et al., 2014; Zook, Graham, Shelton, & Gorman, 2010), having access to a diverse set of people with “web literacy”. In other words, we make no claim that our sample of workers from each country represents the general population of that country.

The first step involved crowdsourcing a simulated search task (Borlund, 2000) to participants in four European locations (Great Britain (GB), Germany (DE), Spain (ES) and Italy (IT)). Queries were cleaned and aggregated, and a content analysis was performed, resulting in a taxonomy of themes characterising participants’ information needs. In the second step, we used the aggregated queries to study the similarity of results retrieved by Google. Images were collected, and then content analysed by an image tagging tool. Finally, we conducted three analyses to understand the similarities and differences in the “view” on the pandemic portrayed: (i) we compared the image overlap in terms of the URLs retrieved, (ii) we analysed the content of the images retrieved (i.e., the respective content tags), (iii) we performed a *localisation analysis*, to understand the proportion of images presented to a user, which are sourced from a domain located in the user’s respective country. We also investigate how aspects (such as the thematic category of query or query language) influence the similarity of results.

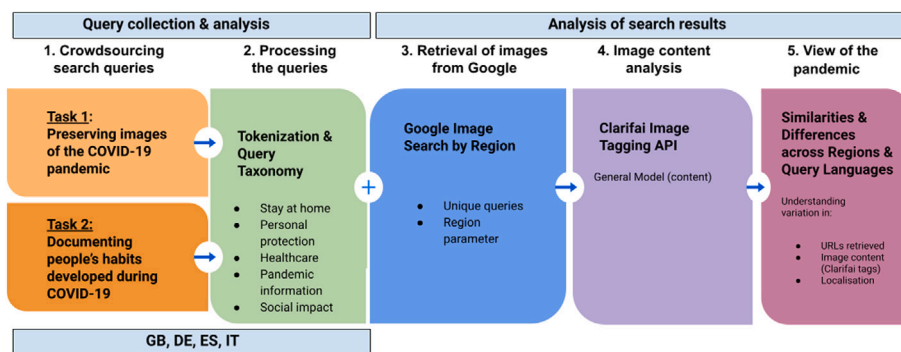


Fig. 1. Overview of the data collection and analyses in the two-step approach.

3.2. Crowdsourcing search queries (Step 1)

We required a crowdsourcing platform with a large pool of workers established in Europe. The Clickworker⁵ platform advertises an attractive population of workers, with 30% being located in Europe. Additionally, it features a function for pre-selecting eligible workers based on country of residence and gender. To test the platform's claims and whether we could achieve the desirable distribution of demographic characteristics in the sample, we performed a test run targeting the four countries. Through this process, we also estimated the time required to complete the task, which was ten minutes. Following the recommendation of the platform⁶ we rewarded workers with €1.60 per completed task according to the above estimation.

We then executed four crowdsourcing “campaigns”, one for each target country, in which we sought responses from 50 men and 50 women, for a total of 400 participants. Our task was set up as a questionnaire using the template provided by the platform. The task, described in detail below, was presented in English to workers across locations, to ensure uniformity. However, workers were encouraged to complete the task in the language that they usually search the Web (a.k.a., we included the following in the Instruction to the workers: “Please provide your queries in the language you usually search the Web.”). Furthermore, we asked workers to state explicitly the language of their queries. The crowdsourcing tasks for all four locations were posted in parallel and run for a couple of days during mid-September 2020.

3.2.1. Description of simulated search task

Crowdworkers selecting our task were first presented with the project information sheet, which: (i) identified our research group; (ii) provided information on how the data would be used, and assured participants that responses were anonymous; (iii) asked participants for informed consent. They were then provided with two prompts, and asked to provide three search queries (of up to five words per query) in response to each prompt. They were also told that “you may test your queries in Google Image Search if you wish to check the images retrieved for a given query”. Although we did not require this check, it was encouraged to promote quality responses.

The two prompts were as follows (see also Figure S1 and S2 in the Supplementary Document):

- Prompt 1. The number of photo documentaries that exist depicting the historic pandemic of 1918 is limited. We want to record the Covid-19 pandemic through a photo documentary. Please provide us with three image search queries to search the Web and collect relevant images documenting the current pandemic, and its various dimensions/aspects.
- Prompt 2. We want to record the habits that people developed during the Covid-19 lockdown through a photo documentary. Please provide us with three image search queries to search the Web and collect relevant images documenting these habits. You may include examples of both “beneficial” as well as “harmful” habits.

Prompt 1 aims at collecting queries that give a more general view of the pandemic, also providing rich visual information. For example, an image of a crowd wearing masks may be more valuable than a graph providing statistics about the pandemic, in that it captures what made an impression on people. This is also the reason why Prompt 1 mentions the 1918 pandemic, stimulating the crowdworkers to think about what will be remembered longterm about Covid-19. On the other hand, the goal of Prompt 2 is to collect queries that give a view of the pandemic at the individual level and are more relevant to a person's habits. Thus, “human centric” images are expected to be collected “spontaneously” from the two prompts used in the task.

⁵ www.clickworker.com/clickworker-crowd/

⁶ <https://www.clickworker.com/survey-participants-for-online-surveys/fee-recommendations>

3.3. Processing the queries (Step 2)

3.3.1. Identifying high quality responses

As in similar crowdsourcing tasks, we face the issue of the quality (Gadiraju, Kawase, Dietze, & Demartini, 2015) of the collected results. In this respect, we considered one mitigation method and one auditing method. This helped us establish that the queries used throughout the work are of quality in terms of matching the task requirements and the conceptual requirements of this work (i.e., are as realistic as possible).

As a mitigation method to remove spam, we removed any participant responses that had an overall low quality (i.e., in all six required queries). In the context of our task, low quality responses are ones that provided a link instead of a query, or a reply that is completely out of topic (e.g., queries that addressed the 1918 pandemic). In total, we discarded 50 responses out of the original 400 (see Table S1, for details). This initial data cleaning still allowed us to maintain a fairly balanced dataset in terms of country and gender representation, which was our initial objective (50 men and 50 women per country).

Finally, as a last measure to audit the appropriateness of collected results compared to the objectives of this work, we asked participants to indicate how frequently they used the image search function. 40% of participants self-reported that they use image search daily while 44% reported that they use it 1–3 times per week. Additionally, 25% of participants reported that image search is their principal source of information and for 73% of the participants it is their secondary source. For a detailed report on how participants per country use the image search function, see Figures S3 and S4 in the Supplemental document. These results show that participants are a representative sample of image search users, who frequently use image search as an essential source of information.

3.3.2. Pre-processing the collected queries

The queries from both tasks were merged, and then cleaned and tokenized, creating a gender-balanced dataset for each of the four geographical regions. Steps in the query cleaning were as follows: (i) replacement of tab, newline and multiple space characters with a single space; (ii) all text to lower-case; (iii) all the expressions referring to ‘covid’ e.g. ‘corona’ or ‘coronavirus’, were replaced with the word ‘covid’; (iv) any URLs were removed. We identified all the unique queries collected from the participants in each country and computed the number of appearances for each unique query without considering any duplicates of the same user (i.e. frequency).


3.3.3. Categorisation of the queries

Next, we aimed to categorise the queries in terms of the users’ information needs; thus, we performed a content analysis (Herring, 2009). A manual analysis was preferred over automatic means as to avoid the biases inherent in such approaches, and because our queries, while being natural language texts, are not complete sentences. To this end, an inductive approach was used. Initially, three researchers examined the GB queries, discussing the topics expressed in them, until a consensus on the six required categories was reached. Next, two researchers analysed the remaining queries from the other locations (DE, ES and IT), involving a third researcher to resolve any disagreements. We were careful to consider whether or not additional categories might be needed, given that the six categories were defined through the analysis of the GB data; however, it was found that the taxonomy was applicable across the data from all regions. Each query was mapped onto one and only one category.

The categories defined through the content analysis, are as follows:

- **Stay at home:** Queries affirming or asking about habits or actions while in a stay-at-home restriction or lockdown. These queries describe the habits developed due to the stay at home restriction. Also, it includes queries stating the impact of the pandemic to a person’s mental state and well-being. *Examples:* “covid zoom call”, “covid food delivery”.
- **Personal Protection:** Queries asking or describing a personal protection instruction or measure during the pandemic. If the concept of the query can be interpreted as personal protection measure it is included here; thus, queries about equipment or accessories needed are included as well. This category also hosts queries asking general questions about the “do’s and don’t s” during the pandemic. *Examples:* “face mask”, “hand washing”.
- **Healthcare:** This category hosts queries relevant to the healthcare system, the way it was impacted, and the means/methods for identifying covid-19. *Examples:* “covid vaccine”, “covid test centre”.
- **Pandemic General Information:** General queries regarding the pandemic, e.g., how much it has spread in the world and queries asking for statistical facts. Includes queries asking about covid in certain geographical areas. *Examples:* “covid outbreak”, “covid in Italy”.
- **Society/Community Impact:** Queries asking or describing the impact that the Covid-19 pandemic and measures had in the society and the different communities (i.e., at a collective level). This category includes general queries relevant to social phenomena in time and space that were not present before. *Examples:* “covid empty streets”, “covid NHS clap”.
- **Miscellaneous:** This category is used for any queries that do not fall into any of the above and/or of which the meaning cannot be clearly interpreted. *Examples:* “1918 flu pandemic”, “5 edtech startup”. For the purpose of this study, we do not analyse the queries from this category.

Table 1
Example of Clarifai tags.

Image	General tags
 <p>Q: "supermarket queues" (BBC, 2020)</p>	<p>woman, people, stock, shopping, street, city, adult, shop, man, meeting, group, teamwork, shopping bag, pavement, market, conversation, cart, supermarket, group together, business</p>

3.3.4. Aggregation of the queries

After categorising the queries, we identified that some queries were very similar but contain differences in the word order, synonyms or including the word "covid". To produce a more robust analysis, for each category, we merged similar queries together if they contain: (i) the same words but in different order; (ii) synonyms; (iii) the same words with "covid", "image", or "photo". Queries that are sub-sets of each other, e.g., "covid hospital" and "covid nightingale hospital" were considered to be separate. From each merged group, we selected one representative query, the one with the highest frequency.

3.4. Retrieval of images from Google (Step 3)

We used Zenserp API⁷ to retrieve results from Google Images. To investigate how the results differ when the search is carried out from different locations, we modified the "gl" (geo-location) parameter in the search API. Four different countries are investigated in this study and are represented using their ISO-3166-1 country codes: Great Britain ("GB"), Germany ("DE"), Italy ("IT") and Spain ("ES"). To reduce the variable, we specified the "hl" (website interface) to be "en" (English), and the search engine to be "google.com" throughout the experiment. E.g., a search for the query "face masks" from a user based in Germany is retrieved by specifying the following parameter in Zenserp:

```
params = (('q', 'face masks'), ('tbn', 'isch'),
          ('device', 'desktop'), ('gl', 'DE'),
          ('hl', 'en'), ('search_engine', 'google.com'));
```

We retrieved the top 100 results per query in the period of 18–25 November 2020. For each query, the search request for the four locations were done consecutively (i.e., same day and similar time) to avoid the influence of results volatility (Jimmy, Zuccon, & Demartini, 2018) in the similarity analysis.

3.5. Image content analysis (Step 4)

The top 30 images from each query were tagged using the Clarifai image tagging API.⁸ In recent years, Clarifai has become a popular tool for researchers aiming to understand visual communication behaviours on the large scale (e.g., Arabghalizi, Rahdari, & Brambilla, 2017; Chen, He, & Kan, 2016; Lee, Hoang, & Lim, 2017). Interestingly, Chen and Dredze (2018) recently used Clarifai to understand the content of images shared on Twitter in discussions about vaccines. In a similar vein to previous work, we used the "general" pre-trained Clarifai model that produces a list of twenty keywords that describe the image content. An example is shown in Table 1.

3.6. View of the pandemic (Step 5)

We investigated the variations – the similarities and dissimilarities – of results retrieved by Google Images, based on the aspects of the images retrieved:

1. *Image overlap*: similarity of the source URLs of the retrieved images. We measured the proportion of image overlap (i.e., same image URLs) using the overlap coefficient measure (Vijaymeena & Kavitha, 2016) in the top 30 and the top 100 results.
2. *Concept overlap*: similarity of the content (i.e., tags received from Clarifai) of the retrieved images. The concept overlap is measured using Jaccard similarity of the unique tags, and the cosine similarity of the tags frequency. These measures are computed in the top 30 and top 100 results.

⁷ <https://zenserp.com/>

⁸ <https://www.clarifai.com/>

Table 2

Number of representative queries by category and country. In parenthesis is the total number of queries in each category.

	Stay at Home	Personal Protection	Healthcare	Pandemic General	Society Impact
ES	72 (87)	48 (80)	42 (60)	41 (74)	95 (126)
DE	52 (74)	36 (58)	28 (34)	65 (105)	85 (117)
IT	58 (84)	52 (80)	52 (77)	61 (91)	112 (153)
GB	70 (108)	57 (83)	26 (43)	58 (89)	113 (153)

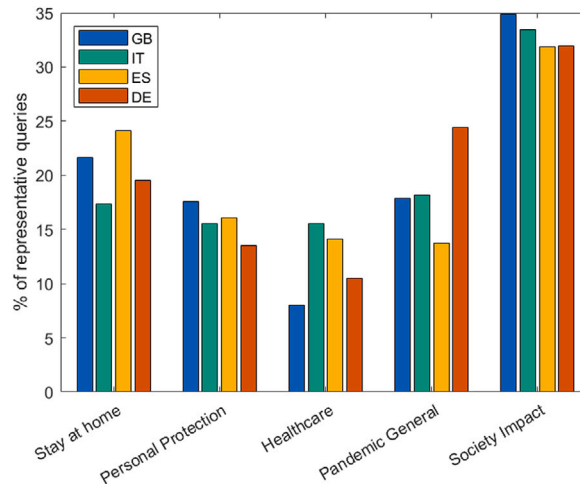


Fig. 2. Percentage of the representative queries per category and country.

3. *Localisation*: the rate of localness of search results (Ballatore et al., 2017), which is quantified using two stages: (i) Identifying the country associated with the domain using suffix identification (e.g., “bbc.co.uk” and “abc.es” are registered in the UK and Spain, respectively), or by retrieving registrant’s country information via the whois protocol. Using this approach, we were able to identify 98% of all the results found in our analyses. (ii) After the domains have been identified, we used a simple, interpretable metric proposed by Ballatore et al. (2017) to quantify the localisation rate of a results set, i.e., the ratio of the local results (i.e., results coming from a domain located in the user’s location) to the number of total results analysed. Thus, the measure ranges from 0 (i.e., all retrieved results are non-local) to 1 (i.e., all results are local with respect to the user). We used the metric to calculate the ratio of localised results in the top 30 and 100 images.

Using these aspects, two analyses were conducted, one in which we varied the country from which the user searched on the queries, and a second in which the language of the query is varied, although the country of access is held constant. We also investigated the influence of thematic categories of queries in the similarity of results.

4. Analysis of user queries

The queries collected from the participants were categorised into one of five thematic categories and then aggregated into similarity groups (i.e., representative queries), as detailed in Section 3.3.4. We compare the distributions of the thematic categories, as they are used across countries, considering the number of representative queries (Table 2).

The categories in Fig. 2 are presented in order, from queries addressing more “personal” search topics (i.e., *Stay at Home*, *Personal Protection*) to more “general” search topics (i.e., *Healthcare*, *Pandemic General Information*) to queries providing a global perspective of the influence of the Covid-19 pandemic to the core of the society (i.e., *Society Impact*). Fig. 2 depicts the percentage of representative queries for each category over the total number of representative queries per country (not including the miscellaneous category). The results presented in Fig. 2 provide an overall comparison of the *variety of queries* collected from each country in each category.

Before going into the analysis of the results we performed a chi-square test of independence to examine the relation between the categories and the country of residence for the set of representative queries. The relation between these variables was significant, $X^2(15, 1223) = 24.0308$, and $p = 0.02$. For this reason we have a closer look to this relationship.

4.1. Analysis of user queries per category

As shown in Fig. 2, in the *Stay at Home* category, the crowdworkers in Spain provided the most representative queries, in contrast to the Italian crowdworkers who reported the lowest variety in comparison to other workers. The German participants reported 4.6%

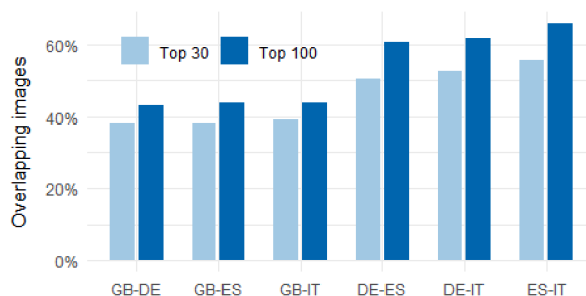


Fig. 3. Comparison of shared images in the top 30 and top 100 (over all 324 queries).

less representative queries compared to the Spanish sample, where as the Great Britain sample had the second highest percentage of representative queries which means that they provided diverse topics of queries in the same category.

Considering the *Personal Protection* theme, we notice that the German sample reported the lowest percentage of representative queries, thus the highest similarity of queries, as shown in Fig. 2, while the Great Britain sample had the opposite behaviour, having the most diverse queries.

In contrast to the rest of the categories, in the *Healthcare* theme, the Great Britain crowdworkers reported less representative queries than in other categories. In contrast, Italian crowdworkers submitted the highest percentage of representative queries in this category which can be justified by the fact that the Italian population had a larger time of exposure than in other countries to the pandemic during the time of the study, and this is reflected in at least in the *Healthcare* category that includes queries relevant to hospitals, symptoms, etc.

Regarding the *Pandemic General Information* category, German crowdworkers have a much larger percentage of representative queries (i.e., highest similarity), at least 6.2% more compared to the rest of the countries. This is the largest difference of representative queries among countries over all the categories, as shown in Fig. 2. It appears that German crowdworkers are “preoccupied” with various general topics that relate to the pandemic and they have a more similar way of expressing those queries compared to workers from other countries. Moreover, the Italian and Great Britain sample of queries have a very similar percentage of representative queries ranking second and third respectively, compared to the other countries.

As per the *Society Impact* category, as shown in Fig. 2, for every country, the percentage of representative queries is the highest among all the categories. This indicates that every sample of queries from each country has a higher focus on describing social phenomenon emerging from the pandemic or the subsequent effects of the measures to restrict the spread of the virus. Queries reported by the Spanish and German workers in this category have almost the same percentage of representative queries and the lowest among all countries. On the other hand, the sample of queries received from GB on this category was the most diverse expanding to different topics (highest percentage of representative queries).

4.2. Overall view of the collected queries

From the observations presented above, it appears that the sample of queries collected from the crowdworkers in Great Britain provides a large diversity of queries for the different categories, as compared to the rest of the countries. In other words, with the exception of the *Healthcare* category, the GB workers had a large percentage of representative queries in all the categories in contrast to the other countries which means that they provide queries of diverse topics in almost all the categories. Thus, in the second part of the study, the analysis of similarity of search engine results, we focus on the GB queries.

5. Analysis of similarity of search engine results

5.1. Analysis of search results of queries accessed from different locations

5.1.1. Queries

We used the GB-sourced queries to investigate how different the results are when the query is executed across locations. 324 representative queries were used in this analysis; the query length ranges between 1–5 words (mean=3.15, SD=0.98). We used four geo-locations (GB, DE, ES and IT) in the search request and retrieved the top 100 image results returned from Google Search.

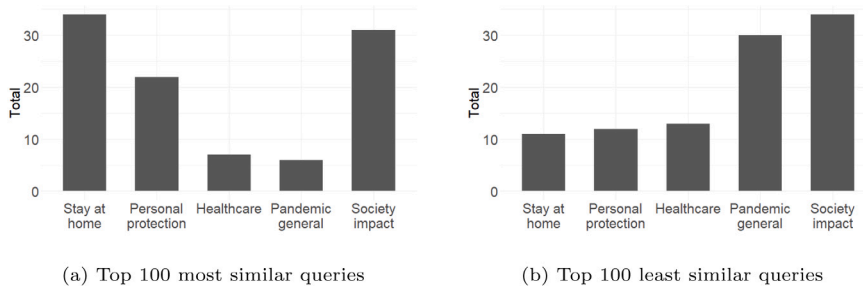
5.1.2. Similarity of results based on image URLs

We first analysed the image overlap in the top 100 results when the same query was accessed by users from different geo-locations. We found that the GB results have around 44% overlap compared to results accessed from DE, ES and IT using the same queries. These proportions are reasonably lower compared to the image overlap between DE-ES, DE-IT and ES-IT results (60%–66%). Furthermore, the average overlap of images in the top 30 (mean=0.46, SD=0.21) was significantly lower than the top 100 (mean=0.53, SD=0.17) over all queries ($t(323) = -19.54, p < 0.05$). The overlap between location pairs are shown in Fig. 3.

Table 3

Mean average image overlap in the top 30 results across all location pairs.

(a) Most similar images	
Query	Overlap
lockdown protest london	94%
face shields philippines	92%
thank you nhs	91%
social distancing schools uk	91%
excessive hand washing	88%
social distancing sign uk	87%
how to get taste back	87%
face mask littering	87%
lockdown baking	86%
home hiit workout	86%
(b) Least similar images	
Query	Overlap
covid 2020	2%
covid social	2%
covid lockdown	2%
covid	2%
covid hospital	4%
covid virus	6%
covid deaths	7%
covid yoga	8%
show me covid infection graphs	9%
covid update	9%

**Fig. 4.** Distribution of categories.

We further found that the overlaps differ significantly between queries (see Table 3). The query “lockdown protest london” retrieved more than 90% of the same images in the top 30, regardless of the location; it achieved a mean similarity of 94% across all location pairs, the highest similarity in our dataset. On the contrary, the query “covid”, “covid social”, “covid lockdown”, and “covid 2020” obtained less than 3% overlap for each location pair (average of 2% overlap over all). This result indicates that with some queries, users may see very different results depending on where they live. In general, results with a more general nature (e.g., those that do not mention the location in the query) are likely to be less similar across regions compared to more specific queries.

Fig. 5(a) shows the average similarity for each category. A one-way analysis of variance shows a significant effect of the thematic category of queries and the similarity of results, $F(4, 319) = 10.16, p < 0.05$. Post-hoc comparisons using the Tukey HSD tests (Tukey, 1949) showed that the least similar category, *Pandemic general information* (mean = 0.33, $SD = 0.16$), was significantly different to three categories: *Personal protection* (mean = 0.51, $SD = 0.22$), *Society/community impact* (mean = 0.45, $SD = 0.20$) and *Stay at home* (mean = 0.54, $SD = 0.19$). We also found statistically significant differences between *Stay at home* queries (the most similar in our study) and two categories: *Society/community impact* and *Healthcare* (mean = 0.41, $SD = 0.21$). The most similar and least similar queries in each category are shown in Table S2 in the Supplementary document.

We further analysed the distribution of categories in two sets: (i) Set 1: 100 queries that resulted in the *most similar images* (averaged across all four geo-locations in the top 30), and (ii) Set 2: 100 queries with the *least similar images*. A chi-square test found a significant difference between the two distributions (Fig. 4) $\chi^2(4, N = 200) = 32.64, p < 0.05$. Queries that produced the most similar results (Set 1) were from *Stay at Home* (34 queries), *Society impact* (31 queries) and *Personal protection* (22 queries) categories. A considerably lower number of *Stay at home* and *Personal protection* queries were found in Set 2 (least similar results), although the number of *Society impact* queries was similar between both sets. Only six *Pandemic general* queries were found in the most similar results, however, up to five times as many queries of this category was found in the least similar results.

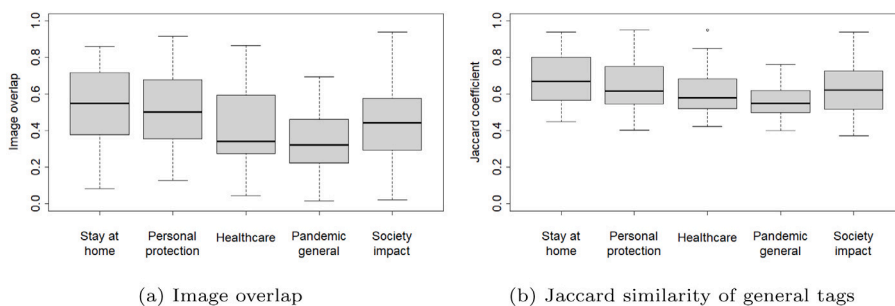


Fig. 5. Similarity of results across categories (Top 30).

5.1.3. Similarity of results based on Clarifai general tags

Our previous results show that the images retrieved for the same query differ considerably based on users' geo-location. In this second comparison, we analysed the similarity of the general tags produced by Clarifai for the top 30 images in each search request. This investigates the similarity between the *concepts portrayed*. We measured the similarity using Jaccard coefficient for the unique tags, and cosine similarity of the term frequency of the tags.

Our results indicates that although the search results have low similarities with regards to the image overlap (mean=0.44), the similarity of concepts portrayed by the general tags are significantly higher (i.e., mean=0.63 (Jaccard) and 0.94 (cosine)). These differences are statistically significant, $t(323) = -31.79$ and -48.71 , respectively ($p < 0.05$). For instance, the query with the least similar images, "covid 2020", shared only < 2% overlap in the top 30 results. However, the corresponding general tags achieved much higher similarity (i.e., Jaccard=0.44 and cosine=0.83). Fig. 5 further shows the similarity across categories for both image overlap and Jaccard similarity of the general tags.

We did not detect any statistically significant differences between categories and the similarity of general tags. However, we found that least similar results were generally retrieved by queries that were more of a general nature, e.g., "covid social", "covid lockdown" and "covid shopping". Interestingly, they often contain the word "covid" (found in 29 out of the 30 least similar queries). In contrast, none of the 30 queries with most similar results contain the word "covid". Half of these queries, however, contain named entities (e.g., "thank you nhs", "uk lockdown breaches", "new york streets during lockdown" and "usa anti mask protests"), which indicate more specific information need. One query, "london standstill during covid", although contains specific location, was one of the least similar queries, i.e., it retrieved different images that portrayed different concepts when accessed from different locations. We investigated this further in the next section.

5.1.4. Results localisation

We analysed whether search localisation was a contributing aspect to the dissimilarity of the search results. In this section, we investigated the rate of local sites found in the search results.

Our findings show that when the queries were searched by a user based in GB, around two-fifths of the top 30 search results were from localised sites, i.e., GB sites ($mean = 0.42, SD = 0.26$). This figure is significantly higher ($p < 0.05$) compared to the localisation rates from other countries: DE ($mean = 0.16, SD = 0.16$), ES ($mean = 0.17, SD = 0.19$) and IT ($mean = 0.11, SD = 0.17$). The localisation rates for these three countries further decreased when considering the top 100 results ($mean = 0.11, 0.10$, and 0.08 , for DE, ES and IT, respectively), while the GB localisation rate remained consistent ($mean = 0.43$).

Interestingly, although the mean rates of localisation were low, we found very high variations between these queries, shown as outliers in Fig. 6. These include general queries such as "covid", "covid patient", "covid 2020", which were heavily localised over all countries. We further found that queries containing the word "covid" has significantly higher localisation in DE, ES and IT countries compared to those without (an average of 0.21 and 0.06, respectively; $p < 0.05$).

Interestingly, we also found specific queries, such as "london standstill during covid", which retrieved localised sites outside GB. E.g., when accessing this query from Germany, many results came from "DW.com", a German site that reports news from Germany and around the world. This brings some insights on why this query retrieved less similar results although it can be considered to be a specific query (i.e., including named entities in the query). The queries with the highest and lowest localised results are listed in Table S3 in the Supplementary Document.

We further found evidence that location-based personalisation greatly influenced the dissimilarity of the results. The average localisation rates in the four countries has strong negative correlation to the overall similarity of results (Pearson's $r=-0.66$). The correlation further increases (Pearson's $r=-0.78$) when taking into account countries outside of GB. We found no correlation between GB localisation rates and similarity of results. These results suggest that location-based personalisation algorithms prioritise local results for some of these queries when accessed from DE, ES and IT; this, therefore, reduced the similarity across the search results.

⁹ NHS stands for National Health Service, which represents the healthcare system in the United Kingdom.

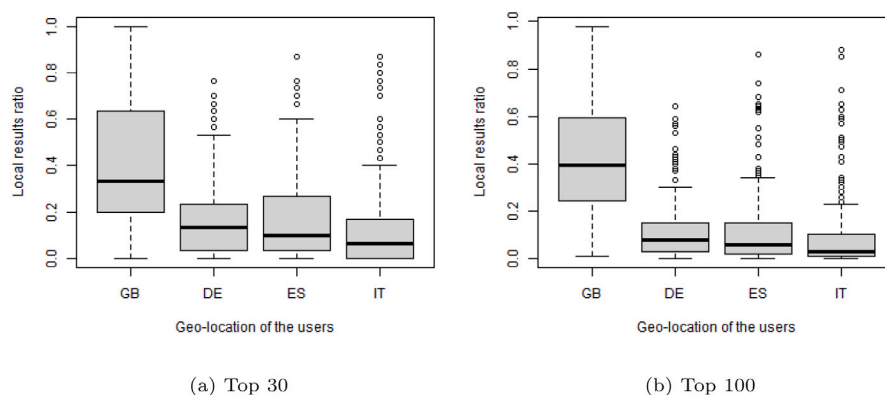


Fig. 6. Rate of localisation in the results (324 English queries).

We further analysed the distribution of country-specific sites in the top 30. Over all 324 queries, results accessed from GB contain <2% results from DE, ES and IT. In contrast, the DE, ES and IT results included around 22%–23% of GB sites. Across the four locations, about one-third of results were from US sites (33%–39% in the top 30, and 37%–38% results in the top 100).

Upon analysing the localisation rates for the top 30 results between the different categories, we found that on average between the four geo-locations, *Stay at Home* and *Personal Protection* queries were localised the least (mean=0.17 and 0.19, respectively). These aligned with the previous findings that showed that the similarity of images in these two categories were the highest. The remaining categories, *Healthcare*, *Pandemic general information*, *Society/community impact* (mean=0.25, 0.25 and 0.24, respectively) were localised significantly higher compared to *Stay at home* queries ($p < 0.05$). However, the differences between the rest of the categories were not significant.

Due to the significant difference between GB results and the other locations, we also analysed results from individual locations separately. As shown in Fig. 7, the localisation rates for all categories in GB results are considerably higher compared to the same categories in DE, ES, and IT results. These differences are statistically significant ($p < 0.05$) for all categories, except for *Pandemic general information* queries between GB and ES results.

We further found a notable difference in the localisation rates across categories in the GB results compared to the other countries. Whilst *Pandemic general information* queries were localised the most for DE, ES and IT results, the localisation rates for these queries for the GB results were the lowest (mean = 0.34, $SD = 0.22$) compared to the other four categories; up to 46% images retrieved were from US sites. The differences between this category and *Healthcare* queries (mean = 0.51, $SD = 0.27$) and *Society/community impact* queries (mean = 0.48, $SD = 0.27$) for the GB results are statistically significant ($p < 0.05$).

Healthcare and *Society/community impact* queries were localised the most (mean=0.51 and 0.48, respectively). Upon further analysis, we found that the *Healthcare* categories contain queries specific to GB, such as “covid nhs workers”, “empty nightingale hospitals”, and “covid testing centre uk”, which might be the reason for the high results from GB sites. We also found many queries in these categories related to the US, such as “usa anti mask protests” and “new york streets during lockdown” which retrieved a majority of US sites in the results. Interestingly, some queries related to other countries, e.g., “photos hospital wards pandemic italy” did not retrieve any IT sites. This suggests that the specificity of a topic to a particular country does not necessarily guarantee a localisation and another aspect (such as language) may also contribute to the localisation rates. We further investigate the multilinguality aspect of the queries in the next section.

5.2. Analysis of search results of multilingual queries from the same location

5.2.1. Queries

This section aims to investigate how different the results are when the queries are executed in languages other than English. For this analysis, we required multilingual queries, i.e., queries that mean the same thing but are written in different languages. Although it was possible to translate the 324 English queries automatically (e.g., using Google Translate), we wanted to ensure that the multilingual queries were realistic and represented queries that native speakers would use. Therefore, we extracted multilingual queries (i.e., queries reported in native language by crowdworkers in Germany, Spain and Italy) that correspond to the original English queries submitted by the GB residents.

To focus the analysis on the effect of multilingualism in the results, we removed queries that appear the same way in different languages, e.g., “covid”, and “pandemia” (which means ‘pandemic’ in Italian and Spanish). In total, we have 50 sets of multilingual queries that appear in at least two languages: 9 queries appear in all four languages, 8 in three languages (i.e., EN and 2 other languages), and 33 in two languages (i.e., EN and another language). Examples are shown in Table S4 in the Supplementary Document.

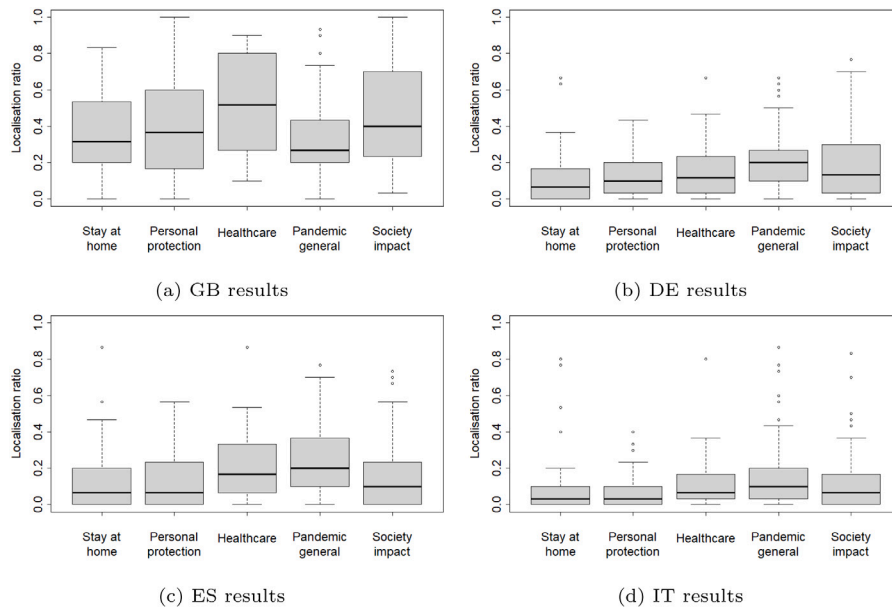


Fig. 7. Rate of localisation per category in the top 30 (324 English queries).

5.2.2. Similarity of results

We analysed the image URLs retrieved by multilingual queries accessed from GB. The results (shown in Table S5) indicate a very low overlap between these queries. Our analysis shows that the average overlap between the top 30 results range between 0.00 to 0.04 (0.00–0.03 in the top 100 results). Queries that retrieved similar results were those that shared similar words, such as (i) DE query: “homeoffice in covid” and EN query: “home office during covid” (sharing 40% similar results), and (ii) DE query: “covid europa” and EN query: “covid europe” (sharing 26.67% similar results). Most of the multilingual queries, however, did not contain many shared words across languages and therefore retrieved very little overlap.

We also analysed the similarity of Clarifai tags for the search results. When accessed from GB, multilingual results achieved Jaccard similarity between 0.35–0.40 and cosine similarity scores of 0.78–0.83, regardless of the language of the queries. These figures are lower than the concept overlap reported in Section 5.1, i.e., when the same queries are requested from different locations. However, we note that the queries used in both analysis have different nature with regards to size, categories, etc. More analysis is therefore required to investigate this aspect in more detail.

5.2.3. Results localisation

Finally, we analysed the localisation of results retrieved by these multilingual queries. Specifying GB as the geo-location, we found a significant difference between the localisation rates of English queries and non-English queries. English queries retrieved an average of 36% localisation rate ($SD = 0.20$). Meanwhile, non-English queries retrieved much lower localisation rate (around 8%–10% GB sites in the results).

Interestingly, when we analysed the results retrieved from other countries (DE, ES and IT), we found that queries written in the official language of the country of residence retrieved much higher localisation rates. The use of German queries (27 queries) in Germany retrieved 79% localisation rate; the use of Spanish queries (21 queries) in Spain retrieved 72% localisation rate. The highest localisation rate in our dataset was found for Italian queries (28 queries) retrieved in Italy, with 93% localisation rate. Similar to the findings reported in Section 5.1, when the language of the queries was different to the official language of the country, very small localisation was found in the results.

6. Discussion

The results reported in Sections 4 and 5 allow us to identify interesting variations in what people see, and to make some specific observations on how Google responds to parameters like the users’ geo-locations and query language. Here, we answer our research questions and relate to findings in previous studies.

6.1. RQ1: what image search queries are of interest to people across regions?

Previous studies (Canchari et al., 2020; Sousa-Pinto et al., 2020) use Google for infodemiology, utilising Google Trends or Baidu Index, in identifying Covid-19 related queries. However, in this work, we collect image search queries through crowdsourcing. This

allows us to have access to a diverse set of people with “web literacy” and collect a wide range of queries of diverse topics, that we would not have the possibility to identify only by looking at popular Covid-19 queries.

Using content analysis, we created a taxonomy of five common themes (*Stay at home*, *Personal protection*, *Healthcare*, *Pandemic general information* and *Society/community impact*) for categorising all user queries from four locations. Across all regions, the queries focused more on the social impact of the pandemic and on new social phenomena experienced during the pandemic, and less on healthcare. The sample of queries collected from the crowdworkers in Great Britain demonstrates a good deal of diversity for almost all categories (except *Healthcare*), compared to the rest of the countries. For that reason, we focus on GB queries for comparing the similarity of Google Images results. Moreover, the five thematic categories of our taxonomy were also used to examine the similarity between image results.

6.2. RQ2: How similar are the results presented to different users?

Based on our findings, when users from different locations search for the same query, only 46% of the retrieved top 30 image results were similar. This figure highlights the high percentage of difference in the visual information provided by Google Images regarding Covid-19.

Analysing the general tags produced by Clarifai image tagger, the concepts portrayed by the different results still achieve medium-high similarity (Jaccard=0.63 in the top 30). These differences were more apparent in the top 30 results than the top 100 results, indicating that users are more likely to see different results to each other if they are only interested in a smaller number of images shown at the top.

We found a low localisation rate when retrieving results from countries other than GB, which indicates that most search results were not highly personalised to match the location of access. The use of English queries might heavily influence these results given the possible lack of relevant documents in non-English sites. However, when considering queries containing the word “*covid*”, which is shared across four countries, the localisation rates were significantly higher. The average localisation rates in DE, ES and IT strongly correlate with the average similarity of results (Pearson’s $r=-0.78$), which suggests that localisation plays an important role in influencing the similarity of results.

When focusing on *multilingual queries*, the similarity of results reduced drastically, often resulting in no similarity between images although the queries were executed from the same location. Although these findings are not surprising given the challenges for retrieving cross-lingual queries, this also indicates that some users (e.g., foreigners or immigrants) who prefer to articulate their queries in their native languages do not have the access to the same information, although they are based in the same country and are affected by the pandemic in the same way.

6.3. RQ3: What aspects influence similarity of results?

Regarding the analysis of the results of image search, we observed that Google image search retrieved results with varying degrees of similarity, both for users who are located in different countries, or those that use queries in different languages. When the same queries were used to retrieve the results from different locations, we identified four aspects that might influence the results.

First, we found evidence that *thematic categories* were represented differently in the results. Our analysis indicated that, in our dataset, *Pandemic general information* queries retrieved the least similar results between the four geo-locations. One-third of the top 100 least similar queries are also from this category. Across the five categories, this category is also the most localised when accessed from outside of GB, which explained the low similarity rate between geo-locations. This suggests that users’ overall view of the pandemic might be fragmented based on where they live. *Stay at home* queries and *Personal protection* queries, on the other hand, are more likely to retrieve similar results regardless of location.

Our analysis further shows that within the same thematic category, the degree of similarity between queries varies widely. This indicates that other aspects might also influence the similarity of results, which brings us to the second aspect, *the specificity of the query*. General queries (e.g., “*covid hospital*”) were more likely to contain different results across locations compared to more specific queries (e.g., “*covid hospital uk*”). This is likely caused by the specific search engine’s (in this case, Google’s) algorithm that prioritises locally-relevant results when processing general queries to filter out irrelevant information and avoid information overload to the users (Bennett et al., 2011; Hannak et al., 2013). For example, the query “*covid hospital*” is highly localised in different countries (0.47 for GB, and between 0.47–0.87 for non-GB), which resulted in 4% similarity across the four locations. Meanwhile, “*covid hospital uk*” resulted in no localisation when accessed from outside GB (0–0.07), and a higher localisation rate in GB results (0.87), achieving 67% similarity across the four locations.

The third factor that might influence the localisation aspect is the *language of the query*. Queries containing shared terms across languages (e.g., “*covid*”) are more likely to be localised across countries, possibly due to the abundance of relevant information from local sites. This results in more variations (i.e., less similarity) for users in different geo-locations. On the other hand, queries that contain terms specifically in English (e.g., “*homeschooling*”, “*lockdown baking*”) might have fewer relevant results from non-GB local sites. Therefore, these types of queries tend to provide users with similar information retrieved from the same (English-language) sites, regardless of users’ geo-locations.

Finally, in addition to the language, we also found high similarity (i.e., image overlap) for *queries describing information needs specific to a location* (e.g., in this case, GB). This includes queries such as “*joe wicks workout*” (0.75), “*clap for carers uk*” (0.64), “*boris johnson sick*” (0.8) and “*lockdown protest london*” (0.94), all of which contained aspects that were specific to GB. Similar to the multilinguality aspect, relevant images to these queries may not be very prominent in non-GB sites, due to the strong connections between these queries to the GB. The rates of localisation across different countries, therefore, are much lower and the similarity between results is higher.

6.4. Impacts of the study

Eskens, Helberger, and Moeller (2017) assert that users have rights to information for truth-finding and other purposes that are relevant to them. During a global pandemic like the one we are currently facing, it is essential that users are directed accurately to the information they need (Makhortykh et al., 2020). However, our study implies that with regards to Covid-19 image search, localisation affects search results and consequently, users do not have access to the same visual information when searching from different countries. Furthermore, similar queries in different languages accessed from GB produced completely different results, which might not be as localised as one would desire; this indicated a different and possibly dangerous treatment of a certain part of the population living in GB.

Whilst previous studies have used Web auditing methods to examine the influence of personalisation and localisation on search results (Kliman-Silver et al., 2015; Lai & Luczak-Roesch, 2019; Robertson et al., 2018; Xing et al., 2014), to the best of our knowledge, no studies have focused on identifying the influence of users' geo-locations in image search results. Moreover, our study also analysed localisation rates in the results, which has not been investigated at this scale before.

Our exploratory study has shed some light on the wide range of difference of results similarity in Google Images. Our findings have shown that search engine users, searching for the same information, may see completely different search results based on where they live and the types of queries they use. The extent to which these results implicate users' information access across regions is an important aspect that we aim to investigate in our future work.

6.5. Limitations of the study

6.5.1. Crowdsourcing task

As discussed in Section 3.2.1, our aim when referring to the lack of documentation through photos during the 1918 pandemic was to "provoke" the crowdworkers to give queries that would generate more human-centric search results. It is possible that our formulation of the task could lead the crowdworkers to draw parallels between the two pandemics that might not have happened naturally. In this respect, to address the risk of introducing bias into our search queries, we removed the queries referring directly to the 1918 pandemic. On the other hand, some queries might not directly refer to the 1918 pandemic but they might have been biased by the wording in Prompt 1. This is a possible limitation to our work, but we believe that the size our dataset and the collective way in which we carried out our analysis have mitigated this form of bias.

In addition, in our attempt to provide a uniform task across all four locations, we posed the crowdsourcing task in English, a lingua franca in Europe. This decision has not affected to a large degree the language in which crowdworkers reported their queries. We received 26.9% English queries from Germany, 18.9% from Spain and 14.6% from Italy. It is also important to consider that participants who are foreign residents of a given country might have replied in their native language.

Finally, another limitation of conducting the crowdsourcing task is that as part of our study, we asked workers to report to us some demographic information as well as to identify how frequently they use image search and how important is image search as a source of information to them. Having participants self-report this information is a technique that has its own limitations (Prior, 2013), but remains the only available tool for crowdsourcing. We remind the reader again that our sample of crowdworkers is not meant to represent the general population of each region, as this was not the aim of the current study.

6.5.2. Analysis of results

We identified a few limitations relating to the similarity analysis of search results. Firstly, previous studies have shown that results seen by logged in users may be different to those accessed using incognito tab (Lai & Luczak-Roesch, 2019; Robertson et al., 2018) due to personalisation, or randomisation (Makhortykh et al., 2020) applied to results shown to the users. In this study, we *did not capture the personalisation aspect* due to the limitation of Zenserp API. This indicates that the images analysed in this study might differ to those shown to real users. However, our findings still assert that, even before personalisation is applied, the results shown to users based in different locations are already significantly different to each other. With personalisation taken into account, it is possible that the differences between results are more significant than those reported in this study.

Secondly, our similarity analysis is based on the overlap of image URLs. Therefore, *duplicate images* with different URLs would have been counted as different results. However, the rate of these is very low and is unlikely to cause major deviation in the results.

Thirdly, due to the small overlap of exact queries between language pairs, our *multilingual analysis was based on a small set of queries*. We plan to gather more multilingual queries in the future to allow us to further assess the impact of thematic categories in multilingual queries.

The last limitation relates to the *quality of the Clarifai general model* used in our image tagging task. Whilst the tagger was useful in identifying general concepts portrayed in the images, it cannot be used to differentiate concepts of varying importance. E.g., with regards to Covid-related images, the overlap of the tag "face mask" across results is more important and relevant than the overlap of tags such as "street" or "sky". However, these varying importance between the different tags are not currently captured, and all concepts are considered to be equally important. We aim to investigate methods to assign weights for these Covid-19 related concepts as future work. Moreover, we found cases where incorrect tags were retrieved, e.g., "drag race" or "rally" were often found for images of people wearing masks, which indicated that the general model was not specifically trained for images related to Covid-19, or a pandemic in general. However, these inaccuracies were constant across the different results and therefore should not cause major differences in the findings.

7. Conclusions and future work

Google search engine plays a key role in the visual information sources that people across the globe access. Our exploratory study aims to investigate how Google presents images of the Covid-19 pandemic, specifically to users in different locations and those using queries in different languages. We used a two step-approach. Firstly, we created a crowdsourcing task to collect a rich and diverse set of Covid-19 related queries for documenting the pandemic from workers in Great Britain, Germany, Italy and Spain. Secondly, we analysed how Google Image Search responds to parameters such as the users' geo-locations and query language in retrieving results for these queries.

An important finding was that Google retrieved different images for all 324 queries, i.e., users in different locations were given at least one different image in the top 30 results. On average, only 46% of images were the same for users across locations. These varied widely between different queries and thematic categories. When considering multilingual queries, users based in the same country retrieved completely different results to each other (less than 4% overlap). These differences are highly influenced by the localisation rates. We further found that English queries accessed in GB retrieved highly localised results compared to results accessed from DE, ES and IT, and that non-English queries have very low localisation in GB results compared to English queries.

Identifying how these different results influence the quality of information that users receive is a challenging yet prominent task that we plan to investigate for our future work. In particular, we plan to develop a method to establish ground truth that we can use to measure the quality of information presented to users. Another direction we wish to examine is the behaviour of crowdworkers in formulating queries based on their cultural background. To this end, we plan to enlarge our dataset, and expand on the idea of how the demographic differences among workers are associated or can explain the provided queries and result differences.

CRedit authorship contribution statement

Monica Lestari Paramita: Methodology, Formal analysis, Investigation, Writing - original draft, Writing - review & editing. **Kalia Orphanou:** Conceptualization, Methodology, Investigation, Writing - original draft, Writing - review & editing. **Evgenia Christoforou:** Conceptualization, Methodology, Investigation, Writing - original draft, Writing - review & editing. **Jahna Otterbacher:** Conceptualization, Supervision, Writing - review & editing. **Frank Hopfgartner:** Supervision, Writing - review & editing.

Acknowledgements

We would like to thank the reviewers for their valuable comments in this work. This project was co-funded by the European Regional Development Fund and the Republic of Cyprus through the Research and Innovation Foundation (Project: EXCELLENCE/0918/0086). This project has also received funding from the European Union's Horizon 2020 Research and Innovation programme under Grant Agreement No. 739578 and 810105 and the Government of the Republic of Cyprus through the Deputy Ministry of Research, Innovation and Digital Policy.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at <https://doi.org/10.1016/j.ipm.2021.102654>.

References

- Arabghalizi, T., Rahdari, B., & Brambilla, M. (2017). Analysis and knowledge extraction from event-related visual content on Instagram. In *Proceedings of the 3rd International Workshop on Knowledge Discovery on the WEB-KD-WEB (vol. 1959)* (pp. 16–27). CEUR-WS.org.
- Ballatore, A., Graham, M., & Sen, S. (2017). Digital hegemonies: The localness of search engine results. *Annals of the American Association of Geographers*, 107(5), 1194–1215. <http://dx.doi.org/10.1080/24694452.2017.1308240>, URL <https://www.tandfonline.com/doi/full/10.1080/24694452.2017.1308240>.
- BBC (2020). Shoppers in Scotland warned of return to supermarket queues. [Online]; URL https://ichef.bbci.co.uk/news/800/cpsprodpb/AAC6/production/_111481734_whatsubject.jpg. (Accessed 28 October 2020).
- Bennett, P. N., Radlinski, F., White, R. W., & Yilmaz, E. (2011). Inferring and using location metadata to personalize web search. In *Proceedings of the 34th international ACM SIGIR conference on research and development in information retrieval* (pp. 135–144). Association for Computing Machinery, <http://dx.doi.org/10.1145/2009916.2009938>.
- Borlund, P. (2000). Experimental components for the evaluation of interactive information retrieval systems. *Journal of Documentation*, 56(1), 71–90.
- Brabham, D. C. (2008). Crowdsourcing as a model for problem solving: An introduction and cases. *Convergence*, 14(1), 75–90.
- Brennen, J. S., Simon, F. M., Howard, P. N., & Nielsen, R. K. (2020). Types, sources, and claims of COVID-19 misinformation. *Reuters Institute*, 1–13.
- Canchari, C. R. A., Chávez-Bustamante, S. G., & Caira-Chuquineyra, B. S. (2020). Exploratory analysis of internet search trends during the COVID-19 outbreak. <http://dx.doi.org/10.36512/RCICS.V31I3.1631.G918>.
- Chen, T., & Dredze, M. (2018). Vaccine images on Twitter: Analysis of what images are shared. *Journal of Medical Internet Research*, 20(4), Article e130.
- Chen, T., He, X., & Kan, M.-Y. (2016). Context-aware image tweet modelling and recommendation. In *Proceedings of the 24th ACM International Conference on Multimedia* (pp. 1018–1027).
- Clay, J. M., & Parker, M. O. (2020). Alcohol use and misuse during the COVID-19 pandemic: A potential public health crisis? *The Lancet Public Health*, 5(5), Article e259.
- Cousins, H. C., Cousins, C. C., Harris, A., & Pasquale, L. R. (2020). Regional infoveillance of COVID-19 case rates: Analysis of search-engine query patterns. *Journal of Medical Internet Research*, 22(7), Article e19483. <http://dx.doi.org/10.2196/19483>.
- Diakopoulos, N. (2015). Algorithmic accountability: Journalistic investigation of computational power structures. *Digital Journalism*, 3(3), 398–415.
- Epstein, R., & Robertson, R. E. (2015). The search engine manipulation effect (SEME) and its possible impact on the outcomes of elections. *Proceedings of the National Academy of Sciences of the United States of America*, 112(33), E4512–E4521. <http://dx.doi.org/10.1073/pnas.1419828112>.

- Eskens, S., Helberger, N., & Moeller, J. (2017). *Challenged by news personalisation: five perspectives on the right to receive information*, no. ID 3060349. Social Science Research Network, URL <https://papers.ssrn.com/abstract=3060349>.
- Gadiraju, U., Kawase, R., Dietze, S., & Demartini, G. (2015). Understanding malicious behavior in crowdsourcing platforms: The case of online surveys. In *Proceedings of the 33rd annual ACM conference on human factors in computing systems* (pp. 1631–1640).
- Haim, M., Arendt, F., & Scherr, S. (2017). Abyss or shelter? On the relevance of web search engines' search results when people google for suicide. *Health Communication*, 32(2), 253–258. <http://dx.doi.org/10.1080/10410236.2015.1113484>.
- Hannak, A., Sapiezynski, P., Molavi Kakhki, A., Krishnamurthy, B., Lazer, D., Mislove, A., et al. (2013). Measuring personalization of web search. In *Proceedings of the 22nd international conference on world wide web* (pp. 527–538). Rio de Janeiro, Brazil: ACM Press, <http://dx.doi.org/10.1145/2488388.2488435>, URL <http://dl.acm.org/citation.cfm?doid=2488388.2488435>.
- Herring, S. C. (2009). Web content analysis: Expanding the paradigm. In *International handbook of internet research* (pp. 233–249). Springer.
- Higgins, T. S., Wu, A. W., Sharma, D., Illing, E. A., Rubel, K., Ting, J. Y., et al. (2020). Correlations of online search engine trends with coronavirus disease (COVID-19) incidence: Infodemiology study. *JMIR Public Health and Surveillance*, 6(2), Article e19702. <http://dx.doi.org/10.2196/19702>, Company: JMIR Public Health and Surveillance Distributor: JMIR Public Health and Surveillance Institution: JMIR Public Health and Surveillance Label: JMIR Public Health and Surveillance Publisher: JMIR Publications Inc., Toronto, Canada. URL <https://publichealth.jmir.org/2020/2/e19702/>.
- Jiang, M. (2014). The business and politics of search engines: A comparative study of Baidu and Google's search results of internet events in China. *New Media & Society*, 16(2), 212–233. <http://dx.doi.org/10.1177/1461444813481196>.
- Jimmy, Zuccon, G., & Demartini, G. (2018). On the volatility of commercial search engines and its impact on information retrieval research. In *The 41st International ACM SIGIR conference on research & development in information retrieval* (pp. 1105–1108). Association for Computing Machinery, <http://dx.doi.org/10.1145/3209978.3210088>.
- Kamoun, A., Maillé, P., & Tuffin, B. (2019). Evaluating the performance and neutrality/bias of search engines. In *Proceedings of the 12th EAI international conference on performance evaluation methodologies and tools* (pp. 103–109). Association for Computing Machinery, <http://dx.doi.org/10.1145/3306309.3306325>.
- Kliman-Silver, C., Hannak, A., Lazer, D., Wilson, C., & Mislove, A. (2015). Location, location, location: The impact of geolocation on web search personalization. In *Proceedings of the 2015 ACM conference on internet measurement conference* (pp. 121–127). Tokyo, Japan: ACM Press, <http://dx.doi.org/10.1145/2815675.2815714>, URL <http://dl.acm.org/citation.cfm?doid=2815675.2815714>.
- Koenig, A. (2020). The algorithms know me and I know them: Using student journals to uncover algorithmic literacy awareness. *Computers and Composition*, 58, Article 102611.
- Lai, C., & Luczak-Roesch, M. (2019). You can't see what you can't see: Experimental evidence for how much relevant information may be missed due to Google's web search personalisation. In I. Weber, K. M. Darwish, C. Wagner, E. Zagheni, L. Nelson, S. Aref, & F. Flöck (Eds.), *Lecture Notes in Computer Science, Social Informatics* (pp. 253–266). Springer International Publishing, http://dx.doi.org/10.1007/978-3-030-34971-4_17.
- Lee, R. K.-W., Hoang, T.-A., & Lim, E.-P. (2017). On analyzing user topic-specific platform preferences across multiple social media sites. In *Proceedings of the 26th International Conference on World Wide Web* (pp. 1351–1359).
- Makhortykh, M., Urman, A., & Roberto, U. (2020). How search engines disseminate information about COVID-19 and why they should do better. In *The harvard kennedy school (HKS) misinformation review* (vol. 1).
- Messaris, P., & Abraham, L. (2001). The role of images in framing news stories. In *Framing public life: Perspectives on media and our understanding of the social world* (pp. 215–226).
- Monzer, C., Moeller, J., Helberger, N., & Eskens, S. (2020). User perspectives on the news personalisation process: Agency, trust and utility as building blocks. *Digital Journalism*, 8(9), 1142–1162. <http://dx.doi.org/10.1080/21670811.2020.1773291>.
- Novin, A., & Meyers, E. (2017). Making sense of conflicting science information: Exploring bias in the search engine result page. In *Proceedings of the 2017 conference on conference human information interaction and retrieval* (pp. 175–184). Association for Computing Machinery, <http://dx.doi.org/10.1145/3020165.3020185>.
- Pfefferbaum, B., & North, C. S. (2020). Mental health and the Covid-19 pandemic. *New England Journal of Medicine*.
- Powers, E. (2017). My news feed is filtered? Awareness of news personalization among college students. *Digital Journalism*, 5(10), 1315–1335.
- Prior, M. (2013). The challenge of measuring media exposure: Reply to dilliplane, goldman, and mutz. *Political Communication*, 30(4), 620–634. <http://dx.doi.org/10.1080/10584609.2013.819539>.
- Ranard, B. L., Ha, Y. P., Meisel, Z. F., Asch, D. A., Hill, S. S., Becker, L. B., et al. (2014). Crowdsourcing—harnessing the masses to advance health and medicine, a systematic review. *Journal of General Internal Medicine*, 29(1), 187–203.
- Robertson, R. E., Friedland, L., Joseph, K., Lazer, D., Wilson, C., & Jiang, S. (2018). Auditing partisan audience bias within google search. In *Proceedings of the ACM on human-computer interaction* (vol. 2) (pp. 1–22). <http://dx.doi.org/10.1145/3274417>, URL <http://dl.acm.org/citation.cfm?doid=3290265.3274417>.
- Scherr, S., Haim, M., & Arendt, F. (2019). Equal access to online information? Google's suicide-prevention disparities may amplify a global digital divide. *New Media & Society*, 21(3), 562–582. <http://dx.doi.org/10.1177/1461444818801010>.
- Sher, L. (2020). COVID-19, anxiety, sleep disturbances and suicide. *Sleep Medicine*.
- Sousa-Pinto, B., Anto, A., Czarlewski, W., Anto, J. M., Fonseca, J. A., & Bousquet, J. (2020). Assessment of the impact of media coverage on COVID-19-related google trends data: Infodemiology study. *Journal of Medical Internet Research*, 22(8), Article e19611. <http://dx.doi.org/10.2196/19611>, URL <https://www.jmir.org/2020/8/e19611/>.
- Tukey, J. W. (1949). Comparing individual means in the analysis of variance. *Biometrics*, 5(2), 99–114. <http://dx.doi.org/10.2307/3001913>.
- Vijaymeena, M. K., & Kavitha, K. (2016). A survey on similarity measures in text mining. *Machine Learning and Applications: An International Journal*, 3(1), 19–28. <http://dx.doi.org/10.5121/mlaij.2016.3103>.
- Walker, A., Hopkins, C., & Surda, P. (2020). Use of Google Trends to investigate loss-of-smell-related searches during the COVID-19 outbreak. *International Forum of Allergy & Rhinology*, 10(7), 839–847. <http://dx.doi.org/10.1002/alar.22580>, eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/alar.22580>, URL <https://onlinelibrary.wiley.com/doi/abs/10.1002/alar.22580>.
- Xing, X., Meng, W., Doozan, D., Feamster, N., Lee, W., & Snoeren, A. C. (2014). Exposing inconsistent Web search results with Bobble. In M. Faloutsos, & A. Kuzmanovic (Eds.), *Lecture notes in computer science: vol. 8362, Passive and active measurement* (pp. 131–140). Springer International Publishing, http://dx.doi.org/10.1007/978-3-319-04918-2_13, URL http://link.springer.com/10.1007/978-3-319-04918-2_13.
- Zook, M., Graham, M., Shelton, T., & Gorman, S. (2010). Volunteered geographic information and crowdsourcing disaster relief: A case study of the Haitian earthquake. *World Medical & Health Policy*, 2(2), 7–33.