4 July - 8 July 2022

According to Google:

Visual epistemologies of climate change and biodiversity loss

SUMMARY AND FINDINGS

The Google Images search engine is probably the most important online gatekeeper of visual culture worldwide, with over two billion searches conducted on the platform every day. Despite the platform's importance, there is a scarcity of analyses on how Google determines the visibility of images within its search rankings. For example, is "authority" determined in a similar way to its textual web search results? Or do computer vision techniques offer new criteria for promoting images and the websites that host them?

This research project focused on the visualisation of [climate change] and [biodiversity loss] by building a dataset of country-specific search rankings from Google Images. We scraped data from six different countries, selected for their political and geographic diversity in relation to the environment: Australia, Netherlands, Nigeria, China, Mexico and Brazil. We built on findings from a 2019 DMI Summer School project, Climate Image Spaces, that suggested Google Images shows a dehumanised visual vernacular of climate change dominated by landscapes, charts and the Earth.

Google Images has a homogenising effect on the visual vernaculars of both climate change and biodiversity loss. For climate change, generic representations of climate change dominate. In particular, a small number of "iconic" stock images are highly ranked across the different countries. Many of these images have a half and half "left/right" layout suggesting the current, often dystopian, situation on the left, with a more hopeful utopian future on the right (occasionally this implied temporality is reversed). Examples of these images include "hand in earth", "landscape" and "tree". For biodiversity loss, there is also homogeneity, although with some greater country-specific diversity. Stock imagery is less dominant, but half and half images remain important, such as 'forest as lungs'. Although biodiversity loss covers many different aspects of environmental degradation, the imagery focuses almost exclusively on deforestation. Images appear to be ranked on whether they "accord" with Google's envisioning of climate change as time-less, place-less, human-less and cause-less. Websites not according to Google's visions of climate change and biodiversity loss are unlikely to be ranked highly by Google Images.

MAIN RESEARCH QUESTION

How does Google Images see and show climate change and biodiversity loss?

RESEARCH SUB-QUESTIONS

To what extent does Google Images homogenise or diversify visions of climate change and biodiversity loss?

What do these results tell us about Google Images' ranking regime?

Data collection

We used a Python script and VPNs to collect the top 50 images associated with [biodiversity loss] and [climate change] on Google Images for six countries: China, Brazil, Netherlands, Nigeria, Mexico. This resulted in a corpus of 300 images for each search term. We identified the most similar/mobile images across datasets using Google Cloud Vision, Clarifai and Memespector. Using vision APIs we identified labels and web entities associated with the images, which we used to gain a deeper understanding of the logic underpinning the sorting of images. Manual coding was then used to identify specific types of images which were not characterised by Google or Clarifai labels (e.g. half and half images, degraded forests). Using semiotics to inform our analysis manual coding served to differentiate between identical and similar images. Country specific datasets were then merged into a master spreadsheet for each term and used to identify top 10 images.

We repeated the same process and expanded our data collection using the Search Engine Scraper developed by the DMI to scrape the top 10 ranking websites of Google. We extracted web domains and used the Triangulate tool to compare Google's Search and Images results in each country.

Data curation

Data and protocols were collated in a shared Google Drive folder. A range of tools were used for data collection, in particular a Python script and DownThemAll, and for data analysis: Clarifai, Cloud Vision, Memespector, Imagesorter, Rawgraph and Gephi. Data was centralised in csv format using Google Sheets. We circulated around the images, displaying them in different ways: according to the initial Google Image ranking (as on the image wall), using most frequent labels and web entities, and also by colours.

Google Cloud Vision API web detection feature draws and updates its information in relation to the authority of landing pages an image is attached to. By default, Google limits the range value to ten pages containing full matching images (this can be changed in the "maxResults" settings). Both clicks on the images and query logs that match image captions on sites that Google prioritises seem to influence the selection of 10 matching pages and web entities (or references that derive from the site-specific textual environments of the image).

Visualisation and Analysis

We used image walls to visually analyse search rankings, identifying patterns and some repeated images, highlighted by recurring image types; images illustrative of these recurring types are displayed for each search term. Deforestation is the only source of biodiversity loss represented in the images, locating the issue as something taking place predominantly in developing countries, rather than globally. Degraded forests feature in around half of the top 50 images for Nigeria, Mexico and Australia but are much lower in the Netherlands and absent for China. In the latter two countries, scientific images are more prevalent. Contrary to calls to consider climate change and biodiversity loss as interrelated, the issues appear largely separate on Google Images, both in the images presented and the 'web entities' that Google uses as references.

In the alluvial visualisation, we focus on the three most frequently appearing stock images for each query, showing where these images are found across the web. Our manual coding shows that the climate change stock image space is distributed to a considerably higher extent, though images frequently appear on news websites. "Religion" and "governmental sites" are specific for the climate change style space. Though biodiversity loss images remain mostly attached to stock image websites, they share sites with climate change images in the category "research and education".

Stock image spaces and their multi-situatedness:

Where else on the web can we identify matching images? Which web sites dominate a given style space according to Google?

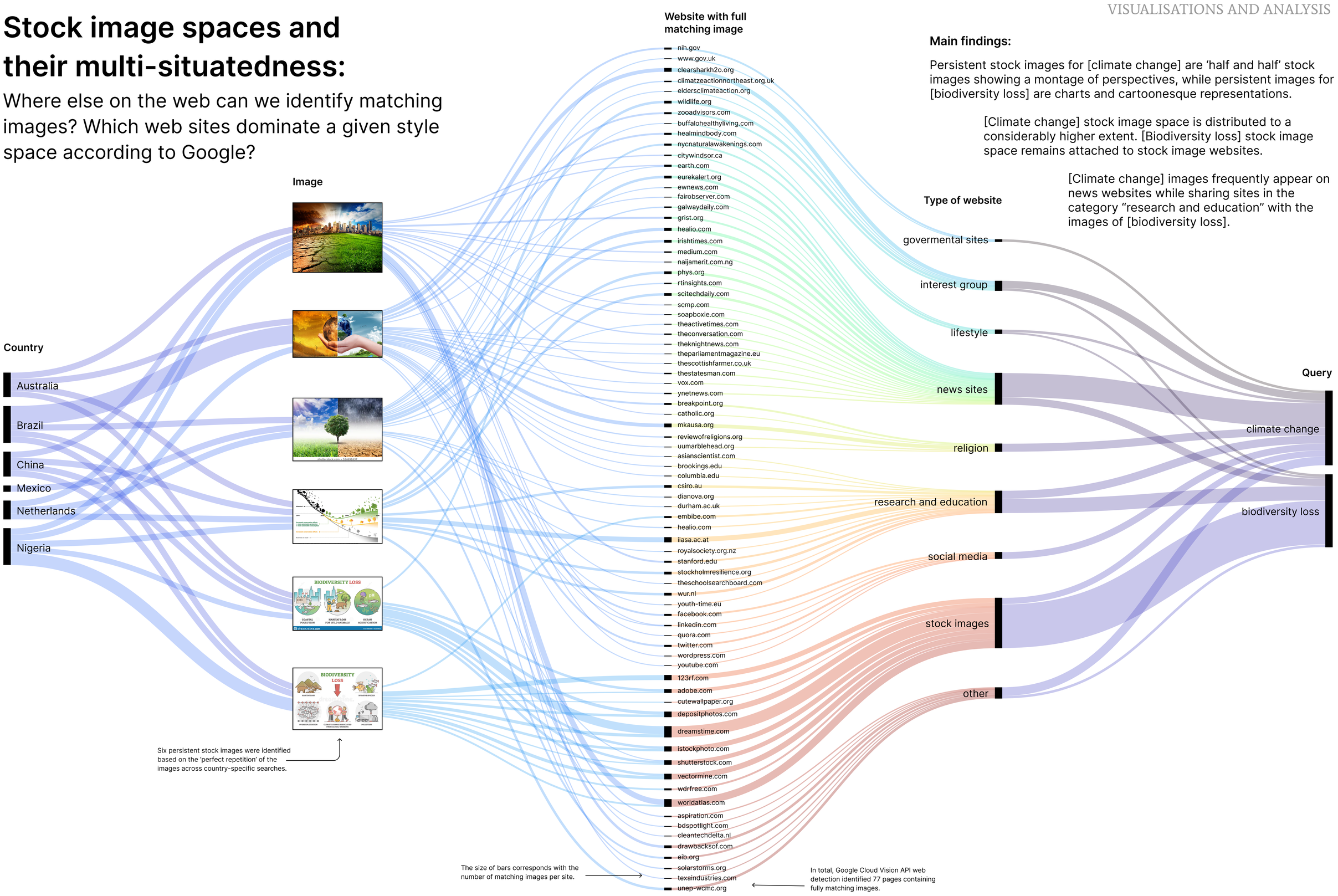
VISUALISATIONS AND ANALYSIS

Main findings:

Persistent stock images for [climate change] are 'half and half' stock images showing a montage of perspectives, while persistent images for [biodiversity loss] are charts and cartoonesque representations.

[Climate change] stock image space is distributed to a considerably higher extent. [Biodiversity loss] stock image space remains attached to stock image websites.

[Climate change] images frequently appear on news websites while sharing sites in the category "research and education" with the images of [biodiversity loss].



PARTICIPANTS

Maud Borie, Laura Bruschi, Ariel Chen, Daniele Dell'Orto, Matthew Hanchard, Warren Pearce, Elena Pilipets, Alessandro Quets, Zijing Xu

Scan to get to the wiki page of the project:

https://wiki.digitalmethods.net/Dmi/AccordingToGoogleImages