

Wikipedia, a source of information on natural disasters biased towards rich countries

February 6 2020



Credit: CC0 Public Domain

Floods are the natural disaster that cause the most damage each year throughout the world. Valerio Lorini (JRC-UPF), Javier Rando (UPF), Diego Saez-Trumper (Wikimedia), and Carlos Castillo (UPF) are the authors of a study they are to present at the 17th International Conference on Information Systems for Crisis Response and Management (ISCRAM 2020), Virginia Tech in Blacksburg, Virginia (USA), from 24 to 27 May, entitled: "Uneven Coverage of Natural Disasters in Wikipedia: the Case of Floods."

The study corresponds to a line of research led by Carlos Castillo, coordinator of the Web Science and Social Computing group (WSSC) at the Department of Information and Communication Technologies (DTIC), UPF, within the active collaboration it enjoys with the Joint Research Center (JRC), the body that advises the European Commission on scientific and technical issues. The principal investigator is Valerio Lorini (JRC-UPF), a student of the Ph.D. programme in ICT at UPF who is being supervised by Carlos Castillo, with Javier Rando, co-author and student of the UPF bachelor's degree in Mathematical Engineering in Data Science.

In the management of [natural disasters](#), access to unofficial data offers the opportunity to dispose of different information from that available through other means. It can also serve to detect bias in news content. "We believe that Wikipedia is a valuable, free source of information and that it could be beneficial to researchers working on reducing the risk of disasters if the biases are identified, measured and mitigated," Castillo asserts.

In their study, the authors focused on the English version of Wikipedia, which they considered by far the most complete version of this encyclopaedia. Wikipedia, an encyclopaedia that is produced collaboratively, contains detailed information on many natural and human disasters, especially when incidents result in a large number of

casualties, and its editors are particularly adept at adding real-time information, as the crisis develops.

As a source of information related to natural disasters, the authors show that on Wikipedia, there is a greater tendency to cover events in [wealthy countries](#) than in poor countries. By performing careful, large-scale analysis of automatic content, "we show how [flood](#) coverage in Wikipedia leans towards wealthy, English-speaking countries, particularly the USA and Canada," they claim in their work. "We also note that the coverage of flooding in low-income countries and in countries in South America, is substantially less than the coverage of flooding in middle-income countries," they add.

For this research the authors estimated the coverage of floods in Wikipedia taking many variables into account: [gross domestic product](#) (GDP), gross national income (GNI), geographical location, the number of English speakers, fatalities and various indices describing the country's level of vulnerability.

They have identified a set of reliable references about floods

With the support of hydrologists, one of the contributions of this work is a set of validated references from several independent organizations that collect data on floods for different purposes: insurers, government agencies, the UN, etc. They all collect data on flooding on a global scale and dispose of reliable databases to work with and compare.

Having identified the sources of information, the authors moved to the experimental phase of the study. Using 458 events that had been reliably described as floods, according to the records of two or three sources of reliable data: Europe's Floodlist; the United Nations' Emergency Events

Database (EM-DAT), and the Dartmouth Flood Observatory (DFO) of the University of California (USA), the authors compared these data with the entries in Wikipedia to locate these events and see if they were consistent or not with the data sources contrasted in terms of location and time references.

"The results of our analysis are consistent over several dimensions, and draw a box where Wikipedia coverage is biased towards some countries, particularly the most industrialized and where large settlements are English speaking, and at the expense of other countries, particularly lower income, more vulnerable ones," the authors suggest.

The results show that the tools that use data from social networks or collaborative platforms should be carefully evaluated to avoid bias, and that Wikipedia editors must make a greater effort to cover [disasters](#) suffered by the neediest countries. These results correspond only to one possible type of natural disaster, floods, but other types of events could also be considered for study.

More information: Valerio Lorini et al. Uneven Coverage of Natural Disasters in Wikipedia: the Case of Flood. arXiv:2001.08810 [cs.IR]. arxiv.org/abs/2001.08810

Provided by Universitat Pompeu Fabra - Barcelona

Citation: Wikipedia, a source of information on natural disasters biased towards rich countries (2020, February 6) retrieved 19 April 2024 from <https://phys.org/news/2020-02-wikipedia-source-natural-disasters-biased.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.