

project, Badr Jaidi and his team, the Social Landslides group, trained computers to automatically extract useful information from relevant news articles about landslides that were posted to Reddit. In this Q&A, he discusses how this tool could end up saving lives.

Why do we need this tool?

According to the World Health Organization, landslides are more widespread than any other geological event. They're so destructive, and we don't have that much data about them. The more accurate landslide data you have, the more it's possible to accurately predict which places have higher risk, which could ultimately save lives.

NASA collects such information in a public database called the Cooperative Open Online Repository, or COOLR, and uses this to predict when and where landslides will occur. But people have had to manually submit landslide information or search for news articles and data one by one, which is pretty tedious. Our tool automates that process, completing in minutes what previously might have taken months.

That would free up resources for more important research, and would also mean we get more data, faster, potentially improving research in landslides generally, as well as NASA's landslide predictions.

How does it work?

Guided by BGC Engineering Inc. and NASA for our capstone project, our team designed a tool that scans Reddit for news articles about landslides within a given period of time and then extracts relevant information.

First, a computer model works out whether the article is indeed about

landslides, rather than say, an election where someone wins "by a landslide," or as we also found, articles about Pokémon with earth techniques like "rock slide."

Then, we trained a [natural language](#) processing model on landslide data, teaching it to recognize the information we wanted from an article. This kind of model can understand language, including analyzing sentences. So, we would give it a news article, and ask where a landslide might have happened. The model would predict the answer based on the language involved, for example, "The landslide most likely happened here, according to this sentence," and we would let it know if it was correct or not.

In this way, the computer learns what information to automatically and accurately extract, including when a landslide happened and where, what caused it, and how many fatalities were involved.

This all happens fairly quickly: It returns a month's worth of articles in about 15 minutes, compared with going through them manually to find those pieces of information. The data can then be fed into COOLR. This took us about two months to build. NASA is currently assessing whether the tool can be run as-is or needs some adjustments to use.

Could the tool be used on other social media sites?

We used Reddit because it's free to access their [application programming interface](#) (API). For instance, Twitter's API has a lot of restrictions, and it's quite expensive to access. Also, the amount of data would be enormous.

We wanted to start small and prove it works with Reddit. But it could be expanded to bigger platforms and sources, provided they have news articles. You could even expand the tool to use it for other disasters such

as earthquakes, using the same methodology by training the models with similar datasets.

Improving upon the model and adding more sources from which [landslides](#) can be extracted other than Reddit would ultimately help NASA have more data points, faster. I'll keep my eye on it.

More information: [BGC-NASA Landslide Detection tool](#).

Provided by University of British Columbia

Citation: Students help NASA find landslides by training computers to read Reddit (2022, October 6) retrieved 17 April 2024 from <https://techxplore.com/news/2022-10-students-nasa-landslides-reddit.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.