We're losing the war against hate speech, and someone is doing something about it

Can technology help the fight against the emerging issue of online hate speech? Yes! A new service has been developed together with Civil Rights Defenders to give bigotry a round.

In 2017 and 2018, some 700.000 members of the Rohingya community were fleeing Burma due to ethnic violence, rising from hate speech on Facebook. To counter the ongoing genocide, the $929.43 billion social media giant deployed a few dozen Burmese speaking moderators. Not unexpectedly, the ethnic violence and hate speech continued.

Social media platforms have a significant impact on the world. Through connecting people over the globe, social media has resulted in multiple positive effects. However, the impact is not only beneficial. Social media platforms have also caused polarization and criminal uprisings, leading to events such as the Rohingya genocide in Burma. By tracking the discourse on social media channels, it is possible to get insights into societal movements and warnings about what is to come.

Until today, the tools to get insights into the massive amount of data of social media platforms have been unavailable. The work has been left for initiatives with either significant funds or technical expertise. Many NGOs are trying to help in the fight against online hate speech and bigotry. One of them is the Swedish NGO Civil Rights Defenders. They have done admirable work fighting the underlying issues of hate speech, such as social insecurity, inequality, and poverty. However, the work towards countering hate speech in real-time is difficult. Due to technical boundaries, it just hasn't been possible for them to get insights into the millions of discussions on social media.

Just now, as the social media giants and other initiatives are struggling with a new wave of hate speech emerging from Covid-19, a new platform has been created to aid in the fight against hate speech. Developed together with Civil Rights Defenders, HateHunter is a service that can help NGOs and other smaller initiatives against hate speech by giving them a user-friendly platform that can provide an instant overview of the Twitter discourse on a topic. By using the latest technologies, such as advanced AI and algorithms similar to Google's search engine, the tool can help give insights from thousands of conversations in mere minutes.

Features of the tool include:
- AI analysis of tweets, classifying tweets that contain signs of bigotry, criminal activity and more, and gives automated insights.
- A dashboard overview of the Twitter activity, providing insights into trends, topics, and important users.
- Advanced filtering, letting the user quickly and easily dig into the data.

The tool is available as a web service and can be used by anyone through the open-source Github repository:
https://github.com/luuddan/hate-hunter