
TECH (/CATEGORY/TECH)

Umati: Kenyan platform to fight online hate speech with NLP and Machine Learning in Africa

Emmanuel Amberber (/author/emmanuel) posted on 24th September 2014



How free is the freedom of speech? Freedom can only exist as long as it is not obstructing another's freedom. Freedom of expression is a necessary condition of our civic and democratic society.



Wikipedia (http://en.wikipedia.org/wiki/Hate_speech) defines 'hate speech' as outside the law, speech that attacks a person or group on the basis of attributes such as gender, ethnic origin, religion, race, disability, or sexual orientation.

We must never be indifferent to hate. But the question is

How do you monitor/crawl the digital arena for hate speech?

How do you analyze the speech for how likely it is to stir violence?

How do you find and use non-government ways of countering it?

Introducing Umati

Umati (<http://www.ihub.co.ke/umati>) is a platform that is dealing with the big data of hate speech. As social media is heavily used by Kenyans, and continues to grow in popularity, Umati project has developed the largest database of hate speech from one country to date.

YourStory Africa had reached out to Nanjira Sambuli, Research Manager, iHub, to tell us about the Umati project she is currently leading.



Nanjira Sambuli | Picture courtesy, netzpolitik.org

"The premise for Umati project," says Nanjira, "emerged out of concern that mobile and digital technologies may have played a catalyzing role in Kenyan 2007/08 post-election violence, and that there was seemingly very little being done to monitor the online space in the build up to the 2013 General Elections. The project seeks to better understand the use of dangerous speech in the Kenyan online space, and monitors particular blogs, forums, online

newspapers, Facebook and Twitter. Online content monitored includes tweets, status updates and comments, posts, and blog entries."

The deadly violence that erupted post 2007 elections in Kenya has taken at least 1,200 lives and hate speeches have contributed for such heinous acts greatly. In order not to repeat the same fate during 2013 Kenyan elections, Umati set out to monitor digital dangerous speech that nobody was monitoring at that time. The police was mostly focused reviewing recorded rally tapes and pamphlets, but the team at iHub and Ushahidi wanted to skate where the 'puke' is going to be, since an estimated 19.2% of the country's online population is active on Facebook.



Hate speech is not something Umati can handle alone it needs NGOs and government collaboration to curb it. Nanjira comments, "When it comes to the authorities, their interventions need to go beyond the speech acts we see, and into addressing the deep-seated issues that lead to the spewing of such speech. We don't believe that prosecution is the right approach; it is a very thin line to tread as far as freedom of speech and expressions go. However, through our findings, we have been engaging government bodies such as the National Cohesion and Integration Commission (<http://www.cohesion.or.ke/>) (NCIC) on their approaches towards fostering cohesion and reconciliation in the country."

The Umati process at work

Nanjira explains two processes that the project had gone through. "**Phase 1** of the project entailed a manual data collection and categorization process. Between October 2012 and November 2013, up to 11 monitors scanned a collection of online sites in seven languages: English and Kiswahili (Kenya's official and national languages respectively); Kikuyu, Luhya, Kalenjin, and Luo (vernacular languages from the four largest ethnic groups); Sheng (a pidgin language incorporating Kiswahili, local languages and English); and Somali (spoken by the largest immigrant community)."

"**In Phase 2** (July 2013 onwards), we have been experimenting with automated means around the Umati methodology; we are exploring Machine Learning and Natural Language Processing techniques."

Umati Project has been widened to its second phase and will be adopted in the upcoming elections in Nigeria. This is all thanks to greater use of Machine Learning and Natural Language Processing Techniques.

How does the Umati Project team decide what constitute a hate speech?

A set of speech characteristics (<http://www.ihub.co.ke/blog/?p=19364&preview=true>) collected off the various sites guide what is classified as *dangerous speech*, that is, speech with a potential to catalyze violence, a definition adopted from Professor Susan Benesch's (<http://www.voicesthatpoison.org/>) work.



"The term hate speech does not have a universally agreed-upon definition. It includes, but is not limited to, speech that advocates for or encourages violent acts against a specific group or creates a climate of hate or prejudice that could in turn, encourage the committing of hate crimes. In this context, speech can include any form of expression, including images, film and music. It is important to keep in mind that a hate comment about an individual does not necessarily constitute hate speech, unless it targets the individual as part of a group."

The Umati team has also been observing a self-regulation mechanism in the speech that takes place online, where Kenyan 'netizens' speak up against hateful or speech that aims to incite, by either ridiculing the speaker, interjecting with more positive speech or use of humour/satire to diffuse any tensions. Nanjira gives an example of how this played out in this report (<http://www.ihub.co.ke/blog/wp-content/uploads/2014/06/2013-report-1.pdf>) (pg 34).

She adds, "We applaud this approach and encourage its continuation. What's most fascinating is that this approach has not involved the intervention of any organizations (government or non-government), but is primarily conducted by Kenyan citizens who desire to preserve a healthy online space for engagement and dialogue."

Making sure privacy remains private

Without any online identity, one cannot have a sustained 'online relationship' with other users.

"We only mine publicly available/shared data, which is accessible to all. We typically don't put out the data sets, but do share as requested by other researchers, bound to agreements that such data will not be used or put out in any way that would compromise privacy, especially of individuals as they identify themselves online," says Nanjira.

As more and more people become connected, it brings lots of opportunity and makes ideas spread faster than ever. It also brings huge challenges. At its darkest side, the internet becomes a perfect platform for those who want to spread hate.



What can be done to prevent hate speech?

Nanjira contends that online hate speech is a symptom of a much more complex issue, often from offline socializations and perceptions that precede online interaction. Online conversations, therefore, may be seen to offer a window into the conversations and convictions people have offline, offering a way to better understand what recurring issues need to be addressed. Therefore, any efforts by governments and civil society towards addressing (online) hate speech ought to concentrate on availing avenues for citizens to work out through their misconceptions of other groups. This will foster a deeper sense of a common identity within the country that transcends groupings that are easily manipulated to stir hatred.

She adds, "Instances of hate speech as observed online should not be viewed in isolation, but in the wider context in which they occur. Hardly has hateful speech prevailed in swaying Kenyans towards harmful action, and Kenyans online are fighting for their freedom of expression (which entails a right to offend), in ways that can only be appreciated by looking at it from a broader perspective."

Scaling Umati geographically

Umati will be putting out a framework (including the automated tool they are working on) for online dangerous speech monitoring for others to replicate in their country contexts as they so wish. Reports and findings are readily available on their website (<http://www.ihub.co.ke/umati>). Nanjira adds, "Our methodology has been adopted for a pilot study in Ethiopia (early 2014). We are currently working with local NGOs in Nigeria to assess the viability of running the project ahead of their 2015 elections."

Umati has been supported by various organizations at various steps: Ushahidi, Internews, and MacArthur Foundation. During this phase of the project (till January 2016), Umati will be coming up with open-source tools for conducting the project, and that the innovation will see others adopt it and build upon it.

Where do we draw the line between expressing our mind and expressing hate?



Think before you comment! Think before you share!

[Report an Issue \(/report-an-issue/umati-hate-speech\)](/report-an-issue/umati-hate-speech)

Recommended for you



Author



Emmanuel Amberber

[\(/author/emmanuel\)](/author/emmanuel)

Everything startups, tech, entrepreneurship, and the intersection of India & Africa. You can reach him via emmanuel@yourstory.com

[\(/author/emmanuel\)](/author/emmanuel)