CNGL can help you to automatically identify subtle, non-explicit forms of bullying language that is widespread online but which is typically difficult to detect.

Industry Challenge

Cyberbullying is a growing concern in our increasingly connected society. Much bullying behaviour involves implicit and metaphorical use of language, often including negative stereotypes. Traditionally, however, cyberbullying activities have been detected through basic keyword filtering that can identify only explicitly offensive content.

Recognising the need for a solution that goes beyond keyword-based approaches to detect more subtle forms of cyberbullying, a team of expert researchers and software developers at CNGL have created a platform that can identify offensive language, harassment and negative stereotyping in a social and cultural context.

The CNGL Solution

CNGL’s content analytics and machine learning experts developed a prototype system, ‘Uonevu’, which automatically detects non-literal forms of bullying and negative stereotyping.

Using crowdsourcing, stereotype data is collected via an annotation client. The collected data is then used to train systems to automatically detect non-literal forms of online bullying and stereotyping across multiple languages and various social media channels.

The prototype builds on a set of linguistic features (e.g. the presence of certain phrases or trigger words), classification results (e.g. controversial topics), and stereotype detection. It uses a semantic knowledge base to associate concepts with each other (e.g. “fat people” and “pizza”), which in combination with other linguistic features indicates an instance of cyberbullying. Machine learning then continually improves the cyberbullying identification and classification process.

To learn how your organisation can automatically identify offensive language and cyberbullying, contact: collaboration@cngl.ie

www.cngl.ie