WebPurify

Profanity Filter

CLIENT	TYPE
Code.org	**Non-Profit/Educational**

"In the 6 years following this partnership's start, Webpurify has been catching inappropriate content on a weekly basis - submissions that would otherwise jeopardize childrens' user experience."

Background

Since 2014, WebPurify has worked with Code.org and its Code Studio's Play Lab, a platform for K to 8 courses where students can build stories, dialogues and games online.

Play Lab has a broad reach, appealing to children of many different ages. Because of its non-profit status, it always is looking to be cost-conscious. For these reasons, the non-profit selected WebPurify as its go-to moderation tool to keep shared apps safe for young eyes. The following is Play Lab's profanity filter case study, highlighting its experience with our service:

The Challenge

"When tasked with adding a swearword filter for shared student projects, I began the work of speccing out a homegrown implementation—creating a bad word list, implementing fuzzy matching for those blacklisted words, handling words in non-English locales, storing and updating this list as false positives/negatives were found, etc.," explains Brian Jordan, a software engineer at Code.org. Jordan immediately began thinking that there must be a more effective way to go about this.

The Solution

Play Lab is an open-source project, and the prospect of maintaining a giant list of "bad words" in its repository felt less than ideal. After doing some research, Jordan found WebPurify, and tested it out on some sample data using WebPurify's Integration Console. Since WebPurify's data base and algorithms are constantly updating to catch profanity, the team at Play Lab now had one less thing to worry about in their codebase.

The Result

❝

It worked just as you'd expect! Adding filtering to our website using WebPurify was just a matter of adding a few lines of Ruby code to call and parse the API response. We just pass in the user's locale and run the user's code right through the API"

BRIAN JORDAN

CASE STUDY
Profanity Filter

CLIENT

Code.org

TYPE

Non-Profit/Educational

"In the 6 years following this partnership's start, Webpurify has been catching inappropriate content on a weekly basis - submissions that would otherwise jeopardize childrens' user experience."

The Industry's Leading Profanity Filter, by the Numbers

100B+
Submissions moderated

15+
Languages supported

FILTERS
URLs, email, and phone numbers

CUSTOM
Block/allow label lists

The Solution Cont'd

Play Lab also benefited from the ability to populate its custom "block list" with any additional words it wished to filter that are specific to the project's needs. "We've been consistently impressed with WebPurify's ability to differentiate benign phrases with very similar vulgar ones," Jordan says.

Our Expertise

WebPurify has extensive experience in moderating children's sites, and we take this service very seriously. Our automated profanity filter is designed to catch the creative ways users attempt to submit text with offensive language. We also offer Image Moderation and Video Moderation to protect this sensitive audience from both offensive content and predators.