

ABSTRACT

A method and a system for detecting harmful content on a network are provided. The method comprises: receiving a URL; obtaining, from the URL, a HTML document associated therewith; converting the HTML document into a text; normalizing the text associated with the HTML document, thereby generating a plurality of tokens associated therewith; aggregating, each one of the plurality of tokens into a token vector associated with the HTML document; and applying, one or more classifiers to the token vector associated with the HTML document to determine a likelihood parameter indicative of the URL being associated with the harmful content; in response to the likelihood parameter being equal to or greater than a predetermined likelihood parameter threshold: identifying, the URL as being associated with the harmful content; and storing, the URL in a database of harmful URLs.

Fig. 5