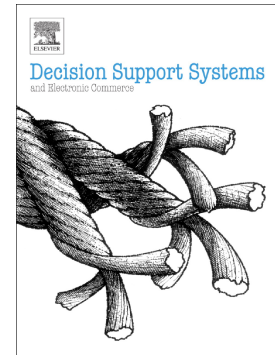


Accepted Manuscript

An Abusive Text Detection System based on Enhanced Abusive and Non-Abusive Word Lists

Ho-Suk Lee, Hong-Rae Lee, Jun-U Park, Yo-Sub Han



PII: S0167-9236(18)30106-4
DOI: [doi:10.1016/j.dss.2018.06.009](https://doi.org/10.1016/j.dss.2018.06.009)
Reference: DECSUP 12967
To appear in: *Decision Support Systems*
Received date: 15 December 2017
Revised date: 25 June 2018
Accepted date: 26 June 2018

Please cite this article as: Ho-Suk Lee, Hong-Rae Lee, Jun-U Park, Yo-Sub Han , An Abusive Text Detection System based on Enhanced Abusive and Non-Abusive Word Lists. Decsup (2018), doi:[10.1016/j.dss.2018.06.009](https://doi.org/10.1016/j.dss.2018.06.009)

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

An Abusive Text Detection System based on Enhanced Abusive and Non-Abusive Word Lists

Ho-Suk Lee, Hong-Rae Lee, Jun-U Park, Yo-Sub Han*

Department of Computer Science, Yonsei University, Seoul 120-749, Republic of Korea

Abstract

Abusive text (indiscriminate slang, abusive language, and profanity) on the Internet is not just a message but rather a tool for very serious and brutal cyber violence. It has become an important problem to devise a method for detecting and preventing abusive text online. However, the intentional obfuscation of words and phrases makes this task very difficult and challenging. We design a decision system that successfully detects (obfuscated) abusive text using an unsupervised learning of abusive words based on word2vec's skip-gram and the cosine similarity. The system also deploys several efficient gadgets for filtering abusive text such as blacklists, n-grams, edit-distance metrics, mixed languages, abbreviations, punctuation, and words with special characters to detect the intentional obfuscation of abusive words. We integrate both an unsupervised learning method and efficient gadgets into a single system that enhances abusive and non-abusive word lists. The integrated decision system based on the enhanced word lists shows a precision of 94.08%, a recall of 80.79%, and an f-score of 86.93% in malicious word detection for news article comments, a precision of 89.97%, a recall of 80.55%,

*Corresponding Author. Tel.:+82-2-2123-5725; Fax: +82-2-365-2579

Email address: {hosuklee, hongraelee, junupark, emmous}@yonsei.ac.kr
(Ho-Suk Lee, Hong-Rae Lee, Jun-U Park, Yo-Sub Han)

Download English Version:

<https://daneshyari.com/en/article/6948314>

Download Persian Version:

<https://daneshyari.com/article/6948314>

[Daneshyari.com](https://daneshyari.com)