

## **Text content analysis for Illicit web pages by using neural networks**

### **Abstract**

Illicit web contents such as pornography, violence, and gambling have greatly polluted the mind of web users especially children and teenagers. Due to the ineffectiveness of some popular web filtering techniques like Uniform Resource Locator (URL) blocking and Platform for Internet Content Selection (PICS) checking against today's dynamic web contents, content based analysis techniques with effective model are highly desired. In this paper, we have proposed a textual content analysis model using entropy term weighting scheme to classify pornography and sex education web pages. We have examined the entropy scheme with two other common term weighting schemes that are TFIDF and Glasgow. Those techniques have been tested with artificial neural network using small class dataset. In this study, we found that our proposed model has achieved better performance in terms accuracy, convergence speed, and stability compared to the other techniques.