Emails are primarily being used for transporting information in a quicker and well-organized way. They are favored in professional as well as personal space because of its attribute of time saving and trading data over huge distances. According to the statistics of the past few years, forgery and fraudulent activities are frequent in emails and these are categorized as ‘spams’. Spam emails utilize time, transmission capacity and storage section; hence it is essential to spot these mails to shield our treasured data and time from being distorted. There are numerous approaches that have been formulated to screen the emails and organize them as spam and non-spam.

The motive of scripting this paper is to analyze recent works done in spam detection and also present a new technique using graylist filter, Boyer Moore string searching algorithm and Naïve Bayes algorithm. Also, observe its working in contrast with traditional Naïve Bayes algorithm.
References


Index Terms

Computer Science  Algorithms
Keywords

Spam, Ham, Tokenization, Classifier.