

{tag}

{/tag}

International Journal of Computer Applications
© 2010 by IJCA Journal

Number 5 - Article 7

Year of Publication: 2010

Authors:

D.Karthika Renuka

Dr.T.Hamsapriya

10.5120/125-241

{bibtex}pxc387241.bib{/bibtex}

Abstract

Unsolicited emails, known as spam, are one of the fast growing and costly problems associated with the Internet today. Among the many proposed solutions, a technique using Bayesian filtering is considered as the most effective weapon against spam. Bayesian filtering works by evaluating the probability of different words appearing in legitimate and spam mails and then classifying them based on that probabilities. Most of the current spam email detection systems use keywords to detect spam emails. These keywords can be written as misspellings eg: baank or bannk instead of bank. Misspellings are changed from time to time and hence spam email detection system needs to constantly update the blacklist to detect spam emails containing misspellings. It's impossible to predict all possible misspellings for a given keyword and add those to the blacklist. In this paper a better and more successful approach for improving E-mail content classification for spam control is proposed. It used the Word Stemming or Word Hashing Technique for improving the efficiency of the content based spam filter. The proposed system extract the base or stem of a misspelled or modified word, to detect spam emails. It considers every misspelled keyword applies a word stemming technique and passes the base word to the content based filter. Using a proposed if-then rule, we can decide whether or not this unknown mail is spam [1]. This paper also provides an Email archiving solution which classifies the E-mail relating to a person, family, corporation, association,

community, or nation.

Reference

[1] Leonard and Hsu, 2001. Bayesian methods: an analysis for statisticians and interdisciplinary researchers. Cambridge University Press, Cambridge.

[2] Bernardo and Smith, 1994. Bayesian theory, John Wiley and Sons, Chi Chester.

[3] Clayton, R. (2004). Stopping spam by extrusion detection. Proceedings of the First Conference on Email and Anti-Spam (CEAS).

[4] Orwant J. et al. Mastering Algorithms with Perl. O'Reilly and Associates, ISBN: 1-56592-398-7, 1999.

[5] Amavisd-new Home Page, <http://www.ijs.si/software/amavisd>, Accessed 01 July 2004.

[6] Send mail Home Page, <http://www.sendmail.org>, Accessed 01, July 2004.

[7] Spam Assassin Home Page, <http://www.spamassassin.org>, Accessed 01, July 2004.

[8] Proc mail Home Page, <http://www.procmail.org>, Accessed 03, Mar 2004.

[9] Graham, P. Better Baysian Filtering. In Proceedings of Spam Conference, 2003.

[10] <http://www.Blog Spam Database.com>

[11] <http://www.Email Spam Filter Word List.com>

[12] <http://www.ceas.cc/papers-2004/172.pdf>.

[13] Internet Users and Spam: What the attitudes and behavior of Internet users can tell us about fighting spam ,Deborah Fallows Pew Internet & American Life Project, Washington, DC, 20036 USA.

Index Terms

Computer Science

Internet Security

Key words

Bayesian Spam Filters

content based spam filter

Word Stemming

Email, Email archiving

