UNIVERSITI TEKNOLOGI MARA

EXPERIMENTAL ANALYSIS ON THE ANTI SPAM EFFECTIVENESS - COMMERCIAL, BAYESIAN AND NGRAM ALGORTIHM

AHMAD KAMAL RAMLI

2006666999

Thesis submitted in partial fulfillment of the requirements for the Degree of Master of Science (Computer Networking)

November 2008

FACULTY OF INFORMATION TECHNOLOGY AND QUANTITAVE SCIENCES
UNIVERSITY TEKNOLOGI MARA
SHAH ALAM

ACKNOWLEDGEMENT

BISMILLAHIRRAHMANIRAHIM

First and foremost, to my creator, ALLAH thanks for giving me the direction and blessing in completing this dissertation.

To Dr Kamaruddin Mamat, my supervisor thanks for being so patients with my attitude in finishing this research project. Although sometime I'm not being fair in meeting and updating you the progress of each milestone. Many thanks to CS778 coordinator and my dissertation examiner, Encik Farok Haji Azmat for making my dreams come true in offering me place in this course and evaluating my works sincerely. Without continues alarm and reminder from you, I might not able to fulfillment this dissertation.

To my friends, Amin, Rafie, Murad, Imran thanks for your continues support in pushing me to complete this dissertation. With continues workload at working place, it is very hard and impossible to furnish my continues commitment towards this research project.

My last compliment goes to my present families for supporting me in everything I did during these critical periods. Aisyah Mat Jasin , thanks for being very supportive and taking care of our son, Abu Bakar, during my absent moments in doing this research project. Ramli Yusuf and Tom Bt Omar , thanks for letting me received sincerest commitment which hard to be carried out by normal father and mother in the world. I'm happy and proud to be your son and this is my tiny rewards to all of your sacrifices towards raising me till present moments.

ABSTRACT

Experimental Analysis On The Anti Spam Effectiveness - Commercial, Bayesian And Ngram Algorithm

With the latest technology and of mail server it exploits the potential of spamming activities to be increase accordingly. People doing spam on they own ways and for their own reasons. Spam is not a virus and it does not contain viruses. Marketing agencies using automated spam structure to perform mass mailing on their promotion and marketing events. At present they are few solutions for tackle spam, using commercial, open source and third party organizations to filter the incoming and outgoing messages. By comparing commercial, Bayesian and N-gram algorithm in rejecting spam and predicting ham messages, it will be a very significant research project for choosing the rights tool to minimize spam activities. By using standard series of text messages which consists of spam and ham words, N-Gram algorithm performed very well. It has the ability to predicting the next alphabet and this is much different with Bayesian algorithm. By applying N-Gram in commercial products, user may receive lots of ham messages inside their inbox. From the test itself Bayesian able to detect only 66.66 % accuracy of ham words inside series of messages. However ,N-Gram score 100% for the same exercise and the algorithm itself have the capability to increase the potential of ham or spam weightage.

TABLE OF CONTENT