

Universiti Teknologi MARA

Spam Image Filtering Algorithm

Muhammad Hazim Bin Shahimi

**Thesis submitted in fulfilment of the requirements for
Bachelor of Computer Science (Hons)
Faculty of Computer and Mathematical Sciences**

January 2015

SUPERVISOR'S APPROVAL

SPAM IMAGE FILTERING ALGORITHM

By

MUHAMMAD HAZIM BIN SHAHIMI

2012509311

This report was prepared under the supervision of the project supervisor, Madam Suraya Binti Masrom. It was submitted to the Faculty of Computer and Mathematical Sciences and was accepted in partial fulfilment of the requirements for the degree of Bachelor of Computer Science (Hons).

Approved by



Madam Suraya Binti Masrom

Project Supervisor

JANUARY 26, 2015

STUDENT'S DECLARATION

I certify that this report and the project to which it refers is the product of my own work and that any idea or quotation from the work of other people, published or otherwise are fully acknowledged in accordance with the standard referring practices of the discipline.



.....
Muhammad Hazim Bin Shahimi

2012509311

JANUARY 26, 2015

ACKNOWLEDGEMENT

Alhamdulillah, praises and thanks to Allah SWT because of His Almighty His utmost blessings, I was able to finish this Final Year Project with the title Spam Image Filtering Algorithm. This project would not have been done without the support and help of many individuals. I would like to extend my sincere thanks to all of them.

Firstly, I would like to give my special thanks of gratitude to my lecturer, Dr Mohamed Imran Bin Mohamed Ariff and also to my supervisor, Madam Masurah Binti Mohamad and Madam Suraya Binti Masrom who had monitored and advised me from the beginning and giving guidance and support to finish this proposal. Their advice and support in completing this Final Year Project report was very helpful for me to finish this work on time. They never gave up in supporting me and also spent a lot of time attending to me even though they are busy with their other works. I am very grateful to have my lecturer and my supervisor.

Furthermore, I would like to thank my parents, family, roommates and friends for their encouragement, cooperation, advices and support from the beginning until now. Everyone in question has contributed by supporting my work, giving ideas and generally being helpful during the progress of my report until it is completed. All the comments, criticism and advices from them are a big help to me. Thank you.

ABSTRACT

Email has become the most important medium of transferring message on the internet. The email users are increasing over the years because it is easy to use and low cost. However, this situation has attracted spammer to advertise their product by sending spam messages to anyone who uses email. As a consequence, the number of spam email has increase unexpectedly. The spam has become a serious problem for the email users. They are flood with a lot of spam in their email. This project is done to help the internet user from being flood by the spam especially spam images. The algorithm in this project will help the existing current spam filtering to enhance the method in filtering spam images. This project applied Optical Character Recognition (OCR) and Bayesian probability to filter the spam images. The evaluation task is done by using formula of Precision, Recall, Error Rate, and Accuracy. Based on the result, the testing dataset has achieved 82% of accuracy. It has shown that the proposed algorithm is good in classifying the images. There are some improvements that can be suggested for future works such as use more data, apply the better OCR technique, use another image feature extraction, and use another classifier.