HOAX EXTERMINATOR

by

Jaidev Soin (BComp)

A dissertation submitted to the
School of Computing
in partial fulfilment of the requirements for the degree of

Bachelor of Computing with Honours

University of Tasmania

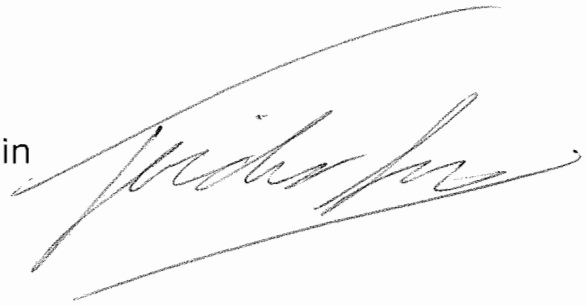
June, 2007

Declaration

I hereby state that this thesis contains no material which has been accepted for the award of any other degree or diploma in any tertiary institution, and that, to my knowledge and belief, the thesis contains no material previously published or written by another person except where due reference is made in the text of the thesis.

Signed

Jaidev Soin



Abstract

This thesis evaluates the potential use of computational techniques to help leverage community knowledge in order to stop hoax emails. Techniques assessed include weighted term based document matching utilising L_p similarity measures and a form of unweighted phrase based matching. These are tested for their effectiveness regarding their ability to identify hoaxes, cope with the evolution of hoaxes and the effects of user added content, and for their low false positive rates. The overall system that these techniques are tested to be a part of supports not only hoax identification, but also user education and feedback. It is observed that the simpler techniques fare the best without generating a large number of false positives, namely phrase based matching and document similarity with an L_1 metric. L_2 measurement is also shown to perform reasonably effectively regarding this application, but suffers from over sensitivity when highly similar hoaxes are compared. A detailed look is taken at hoax emails and the reasons behind their propagation, and a concept system from which all that has been described can be utilised has been outlined. The effectiveness of some of the techniques trialed coupled with a good understanding of both hoax emails and the benefits and limitations that exist when utilising this community content indicate that this fusing of community and technology has great potential.

Acknowledgements

First I would like to thank Chris. You were a pleasure to work with, a fountain of good ideas, and an inspiration to my thoughts. You have certainly changed the way I look at computing, it is no longer just ones and zeros as it now has people at its core as well.

Thankyou to my family. To my father for making my life much easier over these final few weeks. Thankyou Raj for being there for a chat when I needed a break and for rescuing me from my never ending sentences, and thankyou to my mother, for your support and comments such as "You will get it done, I did and I had to write mine on paper!".

A huge thanks to Nate and Ange. Talking to you two had to be my single biggest distraction, but it was also what kept me sane. Nate, knowing you're about is always a comfort and just makes things easier in general. Ange, you have helped me survive my undergrad, and now you have helped me again. You have always made me smile and that always makes life easier and brighter.

Matt, I can't imagine having done Honours without you. Working with you was always fun, and you were always willing and ready to help out. I hope we stay friends for a very long time.

Finally, a thankyou to all those who have helped me enjoy life for the past 12 months both in, and out of uni. Thankyou Neat if you ever read this, thankyou Balmik for giving me someone to best, and thankyou to all those who tried so hard to remember just what it was I was writing my thesis on.

Table of Contents