ABSTRACT

Too much using of internet as communication media, news spreading, and there are a lot of email service provider in internet cause the number of spam email being excessively. It surely can harm the email user because the user have to spend much time to delete spam emails and can cause the storage media on email server being full. Spam email is flooding the internet with many copies of the same message, in a attempt to force the message on people who would not choose to receive it. Spam email usualy consist of commercial message to some product, bussiness message, or even porn message on user who would not want it. At present, there are many spam filtering technique that are developed to force this spam email, for example rule base filtering, naive bayesian filtering and support vector machine. Most of email applications that using spam filtering technique, such as Yahoo Mail, can not understand the semantics of email document, and use a regular expression match, where if a term appears in a particular email, it is filtered. Although this approach is able to filter spam emails, it could occasionally filter some important emails, which might just cotain such term.

This Final Project has designed and implemented a spam email filtering tool using one of Information Retrieval Technique, called Vector Space Model. Vector Space Model act the query as a vector in mutidimensional room. Given an indexing data of spam and legitimate message, so that the spam email filtering tool is able to categorize email, by indentifying content of email to determine which one is spam email .Thus, whenever spam is match, it is filtered.

keyword : spam, email filtering, information retreival, vektor space model.