

ABSTRAK

Banyaknya penggunaan internet sebagai media komunikasi, penyebaran berita serta makin banyaknya layanan penyedia email di internet menyebabkan email *spam* semakin banyak. Hal ini tentu merugikan bagi pengguna email karena harus menghabiskan banyak waktu untuk menghapus email-email spam tersebut dan dapat menyebabkan media penyimpanan pada *email server* menjadi penuh. Email *spam* biasanya berisi pesan komersial tentang suatu produk, usaha, atau bahkan pesan tentang pornografi yang tidak diinginkan oleh user. Saat ini sudah banyak teknik spam filtering yang dibuat untuk mengatasi email *spam* ini, seperti *rule based filtering*, *naïve bayesian filtering* dan *support vector machine*. Kebanyakan dari aplikasi yang menggunakan teknik spam filtering saat ini, seperti *Yahoo Mail* tidak dapat mengenali pola dari dokumen email, dan menggunakan pencocokan ekspresi reguler, dimana jika terdapat suatu kata yang mengandung spam dalam suatu email, email tersebut difilter. Meskipun pendekatan ini dapat memfilter email spam, namun hal ini dapat menyebabkan email-email penting juga difilter karena mengandung term tersebut.

Pada tugas akhir ini telah dirancang dan diimplementasikan suatu perangkat lunak *spam email filtering* menggunakan salah satu pendekatan teknik *information retrieval*, yang disebut *Vector Space Model*. *Vektor Space Model* memperlakukan query sebagai vektor dalam ruang multidimensional. Sekumpulan data indexing berupa email spam dan email legitimate diberikan kepada perangkat lunak *spam email filtering* ini, sehingga dapat mengkategorisasikan *email* dengan mengidentifikasi *content* dari *email* untuk menentukan *email* mana yang merupakan *spam email*.. Sehingga, ketika *spam* tersebut cocok, maka perangkat lunak ini akan memfilternya.

Kata Kunci : spam, email filtering, information retrieval, vektor space model.