

Spam Filter Situs Jejaring Sosial Mahasiswa Menggunakan Regular Expression

Simon Salomon¹, Seng Hansun²

Program Studi Teknik Informatika, Universitas Multimedia Nusantara, Tangerang, Indonesia

Diterima 7 September 2017

Disetujui 20 Desember 2017

Abstract—Spam is an unexpected and unsolicited email sent randomly indiscriminately, directly or indirectly by the sender who has no connection whatsoever with the recipient. The purpose of spam itself is to send information to the recipient, where the content of the sent message generally contains ads that offer non-essential products or illegal products, scams, promotional purposes, or spreading malware designed to hijack computers receiver. Based on the background of the problem, it is necessary anti-spam on a chat or dissemination of information in social networking using regular expression. From this study, the behavioral intention to use at level of 80% means that the user agrees that this website increases user interest in obtaining information and communication, and generates an immersion level of 80% which means the user is very focused when using the website. This website generates value by 98% precision and 98% recall that produce harmonic mean value of 97% so that it can be concluded that it has the precision and recall value harmonious.

Index Terms—social networking, regular expression, spam, website

I. PENDAHULUAN

Perkembangan ilmu pengetahuan dan teknologi khususnya teknologi informasi yang semakin pesat di segala bidang tidak dapat terelakkan lagi. Teknologi informasi merupakan seperangkat alat yang mampu membantu pekerjaan manusia dengan informasi dan melakukan beberapa pekerjaan yang berhubungan dengan proses informasi [1]. Selain itu, teknologi informasi juga memungkinkan sebuah informasi dapat diakses dalam waktu nyata tanpa dibatasi ruang dan waktu.

Namun pemanfaatan Teknologi Informasi belum dimanfaatkan seefektif mungkin. Salah satunya adalah jejaring sosial khusus mahasiswa pada suatu universitas tertentu. Situs jejaring sosial merupakan sebuah situs berbasis pelayanan yang memungkinkan penggunaannya untuk membuat profil, melihat *list* pengguna yang tersedia, serta mengundang atau menerima teman untuk bergabung dalam situs tersebut. Jejaring sosial atau komunitas *online* merupakan konsep pengembangan yang bisa dimanfaatkan di dalam dunia pendidikan. Selain berguna untuk menjalin hubungan antara mahasiswa

di fakultas yang berbeda, juga berguna untuk menunjang dalam meningkatkan efektivitas belajar, dengan cara membentuk komunitas berupa kelompok belajar dan diskusi sesuai dengan bidang dan minat bagi orang-orang yang terlibat di dalamnya. Melalui komunitas *online* ini mahasiswa dapat saling berbagi pengalaman dan informasi mengenai topik yang sedang dibahas tanpa harus saling berhadapan atau saling mengenal satu sama lain.

Aplikasi *web* adalah sebuah kumpulan halaman *web* yang dihasilkan dalam bentuk respon permintaan pengguna [2]. Dengan adanya *website* jejaring sosial ini, diharapkan akan lebih menumbuhkan rasa persatuan dan kekompakan antara mahasiswa, serta menjadi tempat yang tepat bagi mahasiswa untuk dapat mengembangkan kreativitas, ide-ide, berbagi pengalaman serta bertukar pikiran dalam memecahkan suatu masalah, karena *website* dapat diakses kapan saja, di mana saja, dan oleh siapa saja, bukan hanya mahasiswa saja, tetapi juga dosen, maupun staff universitas, tetapi juga masyarakat sekitar, dan bahkan juga dunia. Sehingga akan menjadi sebuah prestasi atau kemajuan yang dapat dibanggakan oleh universitas di mata masyarakat luas.

Spam adalah email yang tidak diharapkan dan tidak diminta, yang dikirim secara acak tidak pandang bulu, langsung atau tidak langsung oleh pengirim yang tidak memiliki hubungan apapun dengan penerima [3]. Tujuan dari spam itu sendiri adalah untuk mengirimkan informasi kepada penerima, dimana konten dari pesan yang dikirim umumnya berisi iklan yang menawarkan produk-produk tidak penting atau produk yang ilegal, umpan untuk skema penipuan, tujuan promosi, atau menyebarkan *malware* yang didesain untuk membajak komputer penerima [4].

Berdasarkan latar belakang masalah tersebut, maka diperlukan *anti-spam* pada suatu *chat* ataupun penyebaran informasi di jejaring sosial dengan menggunakan *regular expression*. Menurut Kuchling [5] sebuah alasan yang sangat bagus untuk menggunakan *regex* adalah karena *regex* sangatlah cepat dan akurat, dimana *regex* menggunakan teknik *finite automata* yang memeriksa setiap karakter pada

string berdasarkan pola yang sudah dirancang sebelumnya. Sehingga, dengan menggunakan *regular expression* maka *anti-spam* akan menjadi lebih efisien karena *regular expression* sangatlah cepat dan akurat dalam mencari dan mencegah *spam*.

II. LANDASAN TEORI

A. Regular Expression

Regular expression atau yang sering disebut sebagai *Regex* menurut Kuchling [5] adalah sebuah formula untuk pencarian pola suatu kalimat atau string. Pada *level* rendah *regex* dapat mencari sebuah penggalan kata. Pada *level* tinggi *regex* mampu melakukan kontrol terhadap data, baik mencari, menghapus dan mengubah data *string*. *Regular Expression* adalah *string* teks yang disandikan secara khusus sebagai pola pencocokan kumpulan *string* [6]. *Regular Expression* dapat diartikan *string* yang memuat kombinasi dari beberapa karakter *normal* dan *metacharacter special* atau *metasequences* [7].

Regular Expression atau yang lebih sering disebut *regex* merupakan sebuah teknik yang digunakan untuk mencocokkan *string* teks, seperti karakter tertentu, kata-kata, atau pola karakter. *Regex* memiliki dua fungsi utama yakni mencari dan mengganti, mencari suatu pola tertentu dalam teks lalu menggantinya menjadi pola yang lain. Sering kali orang beranggapan bahwa *regex* susah dan membingungkan. Namun sebenarnya *regex* sangatlah membantu dalam menemukan pola-pola kalimat, sehingga percobaan terhadap semua kemungkinan pola kalimat tidak perlu dilakukan. *Regular expression* umumnya digunakan oleh banyak pengolah kata atau teks editor dan peralatan lainnya untuk mencari dan memanipulasi kalimat dengan berdasarkan kepada suatu pola tertentu.

Regular Expression digunakan oleh banyak teks editor, *utilities*, dan bahasa pemrograman untuk pencarian dan memanipulasi teks berdasarkan pola. Misalnya, Perl, Ruby dan Tcl memiliki *engine Regular Expression* yang kuat dibangun pada *syntax*.

Fungsi *regex* yang terdapat di PHP untuk mencari *string* menggunakan *Perl-compatible regular expressions*, antara lain: `preg_match()`, `preg_match_all()`, `preg_replace()`, `preg_split()`, `preg_grep()`, dan `preg_quote()`.

B. CodeIgniter

Menurut Basuki [8], *Codeigniter* adalah sebuah *framework* PHP yang dapat membantu mempercepat *developer* dalam pengembangan aplikasi *website* berbasis PHP dibandingkan jika menulis semua kode program dari awal.

Codeigniter menyediakan banyak *library* untuk mengerjakan tugas-tugas yang umumnya ada pada sebuah aplikasi berbasis *web*. Selain itu, struktur dan susunan logis dari *Codeigniter* membuat aplikasi

yang dibuat menjadi semakin teratur dan rapi. Dengan demikian *developer* dapat fokus pada fitur-fitur apa yang dibutuhkan oleh aplikasi dengan membuat kode program seminimal mungkin.

C. Situs Jejaring Sosial

Menurut Firmansyah [9] situs jejaring sosial merupakan sebuah situs berbasis pelayanan yang memungkinkan penggunaannya untuk membuat profil, melihat *list* pengguna yang tersedia, serta mengundang atau menerima teman untuk bergabung dalam situs tersebut. Tampilan dasar situs jejaring sosial ini menampilkan halaman profil pengguna, yang di dalamnya terdiri dari identitas diri dan foto pengguna. Jejaring sosial adalah struktur sosial yang terdiri dari elemen-elemen individual atau organisasi. Setiap situs jejaring sosial memiliki daya tarik yang berbeda. Namun pada dasarnya tujuannya sama yaitu untuk berkomunikasi dengan mudah dan lebih menarik karena ditambah fitur-fitur yang memanjakan penggunaannya. Salah satu contoh situs jejaring sosial yang terkenal adalah Facebook. Dengan beberapa penjelasan di atas dapat diambil kesimpulan bahwa situs jejaring sosial merupakan layanan berbasis *web* dimana digunakan untuk bersosialisasi dan berkomunikasi dengan pihak lain baik dengan teman, keluarga, maupun suatu komunitas yang memiliki tujuan yang sama.

D. Spam

Spam atau *junk email* adalah penyalahgunaan dalam pengiriman berita elektronik untuk menampilkan berita, iklan, dan keperluan lainnya yang mengakibatkan ketidaknyamanan bagi para pengguna *web*. Spam adalah *unsolicited email* yang dikirim ke banyak orang [10]. Spam juga dapat diartikan sebagai *email* yang berisi promosi produk atau jasa, pornografi, virus, dan hal-hal yang tidak penting yang dikirim ke ribuan pengguna.

Sifat-sifat spam menurut Lambert [11] antara lain:

1. Isi atau konten dari *email* tidak relevan dengan minat penerima.
2. Penerima tidak dapat menolak datangnya *email* yang tidak diminta tersebut dengan cara-cara lazim.
3. Pengiriman dan penerimaan pesan tersebut hanya memberikan keuntungan bagi penerimanya.

Spam menurut Raharjo [10] dapat dikategorikan sebagai berikut:

1. *Junk mail*, yaitu *email* yang dikirimkan secara besar-besaran dari suatu perusahaan atau perseorangan yang sifatnya tidak diinginkan.
2. *Non-commercial spam*, misalnya surat berantai atau cerita humor yang dikirimkan secara

massal tanpa tujuan komersial tertentu.

3. *Pornographic spam*, yaitu *email* yang dikirimkan secara massal untuk mengirimkan konten-konten pornografi.
4. *Virus spam*, yaitu *email* yang dikirimkan secara massal dan mengandung virus atau *Trojans*.

Tujuan dari *spam* itu sendiri menurut Cormack dan Lynam [3] adalah untuk mengirimkan informasi kepada penerima, dimana konten dari pesan yang dikirim umumnya berisi iklan yang menawarkan produk-produk tidak penting atau produk yang ilegal, umpam untuk skema penipuan, tujuan promosi, atau menyebarkan *malware* yang didesain untuk membajak komputer penerima. Karena mengirimkan *email* sangat murah, maka pengirim hanya menargetkan sebagian kecil dari penerima, mungkin satu dari seribu atau bahkan lebih sedikit lagi yang dibutuhkan untuk merespon spam yang mengakibatkan keuntungan bagi pengirim. Untuk mencegah tujuan tersebut dapat tercapai, maka dibuatlah *spam filter*.

III. METODOLOGI DAN PERANCANGAN SISTEM

A. Metodologi Penelitian

Metode penelitian yang digunakan dalam perancangan dan pembangunan layanan jejaring sosial ini memiliki beberapa tahapan yang dilalui, yaitu sebagai berikut.

1. Studi Fisibilitas

Tahap ini dapat dilakukan dengan melakukan suatu investigasi awal untuk mendapatkan suatu gambaran terhadap masalah dan menilai secara kasar terhadap ide solusi yang diajukan sebelum menetapkan sumber daya ke proyek [12]. Studi fisibilitas pada penelitian ini dilakukan untuk mengukur kepuasan dari komunikasi dan informasi mahasiswa. Pengukuran ini dilakukan dengan menggunakan kuesioner yang disebar secara *online* maupun melalui media sosial. Target responden dari kuesioner studi fisibilitas ini adalah mahasiswa-mahasiswi dengan jumlah minimal 30 orang. Pencarian dasar teori dari pembuatan kuesioner juga dilakukan pada tahap penelitian ini.

2. Studi Literatur

Dalam studi literatur, pembelajaran terhadap berbagai teori-teori yang berhubungan dengan perancangan dan pembangunan layanan jejaring sosial dilakukan. Teori-teori tersebut antara lain adalah *Regular Expression*, *CodeIgniter*, Situs Jejaring Sosial, Spam, dan lain-lain. Referensi yang digunakan bisa dari berbagai sumber, baik dari *text book* ataupun dari internet.

3. Perancangan dan Pembuatan Sistem

Tahap ini dimulai dengan perancangan struktur

website. Adapun perancangan halaman-halaman dalam *website* ini dibagi atas dua macam yaitu bagi *admin* dan *user* (dosen, *staff*, mahasiswa-mahasiswi, maupun pengunjung biasa). Setelah itu barulah dilakukan perancangan *database*. *Database* digunakan sebagai tempat penyimpanan data yang terstruktur agar dapat diakses dengan cepat dan mudah.

4. Pengujian Sistem

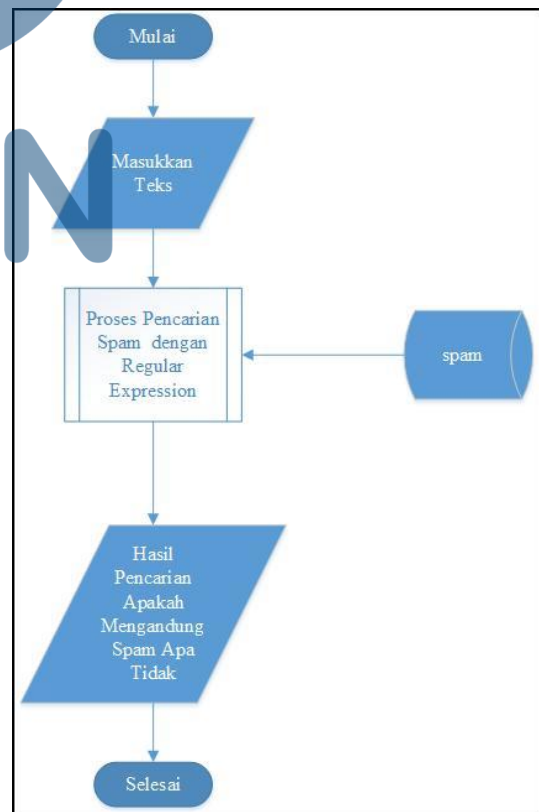
Pengujian sistem dilakukan dengan menggunakan metode studi lapangan, dimana mahasiswa-mahasiswi, dosen, dan karyawan mencoba menggunakan sistem. Metode pengujian dengan studi lapangan dilakukan agar dapat mengukur kegunaan dari sistem yang dibuat. Dalam studi lapangan, responden akan diberikan kuesioner mengenai pengalaman yang dirasakan ketika menggunakan sistem tersebut.

5. Evaluasi

Evaluasi sistem dilakukan dengan menganalisis hasil kuesioner yang diperoleh dari studi lapangan pengujian sistem.

B. Perancangan Sistem

Berikut merupakan proses *Flowchart Filter Spam* dengan menggunakan *Regular Expression* yang dipakai di dalam sistem.



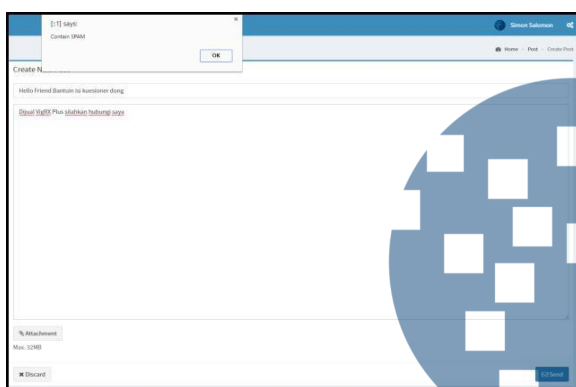
Gambar 1. *Flowchart filter spam* dengan menggunakan *regular expression*

Gambar 1 dimulai dengan *user* memasukkan teks kemudian akan dilakukan pencarian apakah teks tersebut mengandung *spam* apa tidak dengan *regular expression*. Teks yang termasuk *spam* diambil dari *database*.

IV. HASIL IMPLEMENTASI DAN UJICoba

A. Implementasi Antarmuka

Hasil implementasi sistem diperlihatkan pada Gambar 2. Jika *user* menulis kata-kata spam, misal "VigRX Plus" yang berasal dari nama obat kuat maka akan dimunculkan peringatan *spam*. Untuk menutup atau menghilangkan peringatan, *user* bisa menekan tombol "OK". Setelah *user* berhasil menutup peringatan, *user* bisa menulis ulang pesan atau informasi.



Gambar 2. Implementasi Spam Posting

B. Evaluasi dengan Confusion Matrix

Hasil penerapan regex sebagai spam filter pada sistem diujicoba dengan menggunakan confusion matrix. Tabel 1 memperlihatkan hasil confusion matrix dari ujicoba yang dilakukan terhadap 52 *inputan user*.

Tabel 1. Confusion Matrix

| Correct Classification | Classified as | |
|------------------------|---------------|--------------|
| | Spam (+) | Not Spam (-) |
| Spam (+) | 41 | 1 |
| Not Spam (-) | 1 | 9 |

Berdasarkan perhitungan di atas, didapatkan nilai Akurasi sebesar 96%, *Precision* sebesar 98%, dan *Recall* sebesar 98%. Menurut Rijsbergen [13] untuk menyeimbangkan *trade-off* antara presisi dan *recall* dibutuhkan sebuah pengukuran harmonis yang disebut *harmonic mean*. *Harmonic mean* dari *precision* dan *recall* dapat dicapai dengan menghitung nilai *F-measure*. Berdasarkan perhitungan *F-measure* di atas, didapatkan nilai harmoni sebesar 97%. Standar nilai minimum harmoni yang dikemukakan oleh Rijsbergen [13] adalah sebesar 70% sehingga

dapat ditarik kesimpulan bahwa nilai harmoni yang didapat telah memenuhi standar *harmonic mean*.

C. Evaluasi dengan HMSAM

Evaluasi Social Networking Service UMN dengan metode HMSAM dilakukan dengan menyebarkan kuesioner secara *online*.

Tabel 2. Respon Kuesioner HMSAM

| SA | A | N | D | SD | Aspect |
|----|----|----|---|----|-----------------------|
| 21 | 17 | 10 | 2 | 2 | Perceived Ease of Use |
| 19 | 17 | 14 | 1 | 1 | Perceived Usefulness |
| 22 | 16 | 9 | 4 | 1 | Curiosity |
| 17 | 22 | 12 | 0 | 1 | Control |
| 20 | 19 | 8 | 2 | 3 | Joy |

Keterangan:

- SA : Strongly Agree
- A : Agree
- N : Neither
- D : Disagree
- SD : Strongly Disagree

Berdasarkan Tabel 2, dapat dihitung nilai masing-masing aspek dari HMSAM dengan hasil seperti yang disajikan di Tabel 3.

Tabel 3. Hasil Kuesioner

| Aspect | Likert's Score | Interpretation |
|-----------------------|----------------|----------------|
| Perceived Ease of Use | 0,8038 | Strongly Agree |
| Perceived Usefulness | 0,8 | Strongly Agree |
| Curiosity | 0,8077 | Strongly Agree |
| Control | 0,8077 | Strongly Agree |
| Joy | 0,7961 | Agree |

Berdasarkan lima hasil perhitungan di atas, didapatkan hasil yaitu aspek *Perceived of Use*, *Perceived Usefulness*, *Curiosity*, *Control*, dan *Joy* yang disetujui oleh pengguna. Selanjutnya lima aspek ini akan digunakan untuk menghitung nilai dari dua aspek utama, yaitu *Behavioral Intention to Use* dan *Immersion*. *Behavioral Intention to Use* mendapatkan nilai 0,8019 dan *Immersion* mendapatkan nilai 0,8029. Dengan demikian dapat disimpulkan bahwa pengguna sangat setuju untuk terus menggunakan *website* jejaring sosial yang dibangun serta sangat setuju bahwa mereka memiliki fokus penuh saat menggunakan sistem tersebut.

V. SIMPULAN

Layanan jejaring sosial sebagai media interaktif dalam mendapatkan informasi dan komunikasi

mahasiswa dengan *regular expression* untuk mencegah *spam* atau sebagai *filtering spam* telah berhasil dirancang dan dibangun dengan nama *Social Networking Service UMN*. *Website* ini menghasilkan nilai presisi sebesar 98% dan *recall* sebesar 98% yang menghasilkan nilai *harmonic mean* sebesar 97% sehingga dapat disimpulkan bahwa *website* ini memiliki nilai presisi dan *recall* yang harmonis. Berdasarkan perhitungan Likert, *website* ini menghasilkan tingkat *Behavioral intention to use* sebesar 80% yang berarti bahwa pengguna setuju bahwa *website* ini meningkatkan minat pengguna dalam mendapatkan informasi dan komunikasi serta menghasilkan tingkat *immersion* sebesar 80% yang berarti pengguna sangat terfokus ketika menggunakan *website*.

Adapun beberapa saran yang dapat dilakukan terhadap pengembangan layanan jejaring sosial ini adalah sebagai berikut.

1. Mencari dan menambah kata-kata yang termasuk *spam* ke dalam *database*.
2. Tampilan dari aplikasi yang masih terlihat kurang menarik juga bisa diperbaiki agar aplikasi ini menjadi lebih menarik dan interaktif.

3. Menambah fitur video pada *chatting*.

DAFTAR PUSTAKA

- [1] Haag, S and Keen P. 1996. Information Technology, Tomorrow's Advantage Today.
- [2] Murach, Joe; Harris, Ray. 2010. Murach's PHP and MySQL.
- [3] Cormack G.V., Lynam T. 2005. TREC 2005 Spam Track Overview. Canada.
- [4] Cestyakara, A. 2010. Penyaring Spam Bayesian (Bayesian Spam Filter). Dalam Makalah II2092 Probabilitas dan Statistik. Bandung.
- [5] Kuchling, A.M. 2002. Regular Expression HOWTO.
- [6] Fitzgerald, M., 2012. Intoducing Regular Expressions, O'Reilly, Cambridge.
- [7] Stubblebine, Tony. 2007. Regular Expression Pocket Reference, Second Edition. O'Reilly Media, Inc.
- [8] Basuki, Awan. 2010. Membangun Web Berbasis PHP dengan Framework CodeIgniter. Lokommedia: Yogyakarta.
- [9] Firmansyah, Aditya. 2010. Situs Jejaring Sosial Menggunakan Elgg.
- [10] Raharjo. 2006. SPAM. Tersedia dalam: <https://www.cert.or.id/~budi/presentations/SPAM-present.ppt> [Diakses 7 juli 2017].
- [11] Lambert, Anselm. 2003. Analysis of spam, A dissertation in Computer Science at University of Dublin.
- [12] Overton, R. 2007. Feasibility Studies Made Simple. Martin Books Pty Ltd, Boat Harbour, Australia.
- [13] Rijsbergen, C.J. 1979. Information Retrieval. Butterworths.



UMN