

BAB I

PENDAHULUAN

Dalam bab ini dibahas mengenai latar belakang pemilihan topik, permasalahan, ruang lingkup masalah yang membatasi penelitian, tujuan dan manfaat dari penelitian yang dibuat, serta sistematika penulisan yang menjelaskan garis besar penulisan dalam Tugas Akhir ini.

1.1 Latar Belakang Masalah

Di era modern saat ini, internet sudah menjadi kebutuhan yang tidak bisa dipisahkan lagi dari kehidupan manusia. Internet telah membuat revolusi baru dalam dunia komputer dan dunia komunikasi yang tidak pernah diduga sebelumnya. Dengan adanya internet, setiap orang dari berbagai belahan dunia dapat mencari informasi dan berkomunikasi satu sama lain tanpa dibatasi oleh kondisi geografis. Internet memudahkan setiap individu mencari hiburan, belajar, bergaul, dan bahkan berbisnis. Selain itu, internet menyediakan akses 24 jam dengan kecepatan pembaharuan informasi yang tinggi.

Kemudahan mencari informasi dalam internet ini membuat kebutuhan setiap orang terhadap internet semakin meningkat. Dalam *Measuring the Information Society Report* edisi 2014, International Telecommunication Union (ITU) melaporkan bahwa pengguna internet bertambah pesat dengan peningkatan 6,6 persen pada tahun 2014, dimana jumlah saat ini adalah sekitar 3 milyar pengguna. Dengan sedemikian banyaknya pengguna, tampaklah bahwa internet adalah sumber informasi yang paling diminati saat ini.

Kelebihan pencarian informasi melalui internet tersebut, tak dapat dilepaskan dari kekurangannya. Menurut Yoyon (2014), internet tak hanya memberikan kontribusi positif bagi kehidupan tetapi juga ancaman. Terhubung ke internet ibaratnya membuka pintu seluas-luasnya untuk bisa diakses oleh siapapun. Kemudahan mengakses internet membuka gerbang bagi semua jenis informasi, termasuk di dalamnya adalah membuka gerbang bagi konten pornografi dan konten dewasa. Hal ini dikhawatirkan dapat menyebabkan turunnya moralitas para pengguna internet.

Permasalahan ini menjadi semakin pelik dengan adanya hasil penelitian tentang rentang usia pengguna internet. Berdasarkan data dari Asosiasi Penyelenggara Jasa Internet Indonesia, pada akhir tahun 2012, pengguna Internet di Indonesia mencapai 63 juta orang, dimana 60% dari jumlah tersebut berasal dari kelompok umur di bawah 35 tahun. Artinya, sebagian besar pengguna internet saat ini adalah anak-anak, remaja, dan pemuda. Hal ini tampak dari semakin jarangnyanya pemandangan anak-anak yang beraktivitas di luar dan bermain bersama teman. Anak-anak semakin mengenal bermain komputer dan berselancar di dunia internet, tanpa adanya *filter* atau batasan atas informasi yang mereka dapatkan tersebut. Padahal, internet sarat dengan konten dewasa dan pornografi yang akan membuat mereka dewasa sebelum waktunya. Orangtua diperlukan kehadirannya dalam mendampingi anak-anak dalam menggunakan internet, namun kenyataannya orangtua tidak mampu mendampingi mereka senantiasa.

Melihat masalah ini, maka muncullah program-program yang menerapkan *content filtering*. Namun, *content filtering* yang selama ini dikembangkan memiliki sistem yang kaku, dimana apabila satu kata kunci ditemukan, maka website itu langsung diblokir. Akibatnya, berbagai website yang tidak termasuk porno pun terkena blokir. Selain itu, pemerintah Indonesia juga mengembangkan layanan Nawala yang memblokir situs-situs yang termasuk kategori situs porno, situs judi, phishing dan malware. Namun, blokir ini tergantung pada laporan dari pengguna, yang artinya bila tidak ada laporan, maka tidak akan diblokir. Cara ini membuat banyak website porno masih berkeliaran.

Dengan adanya permasalahan tersebut, diperlukan sebuah program yang mampu menganalisa apakah konten yang dibuka oleh seorang anak layak untuk ditampilkan atau tidak. Hal ini dapat diartikan sebagai pentingnya program yang menerapkan *content filtering* berdasarkan isi dari halaman website. Setelah analisa itu dilakukan, diperlukan tindakan langsung untuk dapat mencegah anak-anak melihat halaman website yang terhitung tidak layak ditampilkan. Oleh karena itu, dalam pembuatan program ini diterapkan pengambilan konten halaman website dan perhitungan klasifikasi dengan metode Naive Bayes sehingga diharapkan program ini dapat membantu orangtua membatasi halaman internet yang akan dibaca oleh anak-anak.

1.2 Permasalahan

Dalam tugas akhir ini akan ada beberapa masalah yang harus dihadapi dalam proses pembuatannya. Berikut adalah beberapa masalah yang harus dihadapi dalam penyusunan tugas akhir ini:

1. Bagaimana cara melakukan optimisasi kode program supaya klasifikasi halaman website yang dibuka dapat dilakukan dengan lebih cepat? Permasalahan ini penting karena browser extension dan web service yang dikembangkan ini akan sangat membutuhkan perhitungan yang cepat. Selain itu, metode klasifikasi ini adalah Naive Bayes yang banyak menggunakan operasi perpangkatan, sehingga mahal secara *runtime*.
2. Bagaimana cara mencegah aliran data website terlihat oleh user sementara perhitungan klasifikasi masih dilakukan?
3. Seperti apa cara implementasi Naive Bayes dalam hal *training* dan *testing* agar dapat dipercaya keakuratannya?
4. Apa yang harus dilakukan apabila hasil perhitungan atau hasil perpangkatan Naive Bayes melebihi batas tipe data terbesar?
5. Apa yang harus dilakukan agar program dapat mencegah klasifikasi yang kurang tepat? Misalnya, berita tentang pemerkosaan, seharusnya bukan masuk ke kategori tidak blokir, atau berita tentang posisi dalam berhubungan, seharusnya masuk ke kategori blokir.
6. Bagaimana penyusunan algoritma web service agar dapat menangani pola penulisan kode html yang berbeda-beda dari website-website yang dibuka? Hal ini dikarenakan gaya setiap website berbeda-beda, dan diperlukan algoritma yang tepat supaya pengambilan informasi dapat mencakup sebanyak mungkin website yang ada.
7. Apa saja daftar *stopword* yang diperlukan untuk mendapatkan informasi dari halaman website yang dibuka?
8. Bagaimana caranya agar perhitungan dapat dilakukan terhadap url yang berbeda-beda dalam waktu yang hampir bersamaan?

9. Bagaimana mengatur extension agar secara otomatis dapat melakukan pengecekan terhadap setiap halaman baru yang dibuka, meskipun halaman tersebut baru terbuka di *tab* atau *window* yang belum aktif?
10. Bagaimana cara menghubungkan browser extension dengan web service, serta web service dengan program stemming Sastrawi? Hal ini penting mengingat pemeriksaan terhadap konten hanya dilakukan melalui web service.
11. Apakah algoritma stemming Sastrawi merupakan algoritma yang baik untuk permasalahan ini? Hal ini perlu diperhatikan mengingat algoritma stemming Sastrawi adalah algoritma baru yang merupakan pengembangan dari algoritma stemming lainnya.
12. Apa solusi terbaik untuk klasifikasi website yang secara khusus telah dilabeli blokir atau tidak? Seperti apa cara penyimpanan dan pengambilan datanya?
13. Bagaimana mengembangkan fitur tambahan yang berguna untuk mengembangkan pemblokiran website ini?

1.3 Ruang Lingkup

Ruang Lingkup dalam penelitian ini yaitu:

1. Perangkat lunak yang dibuat dalam tugas akhir ini adalah web service dan extension untuk browser, yaitu Google Chrome.
2. Extension akan mengecek setiap halaman website yang dibuka, dan kemudian akan menghasilkan kesimpulan apakah halaman tersebut diblokir atau tidak. Halaman website akan diblokir bila isinya terhitung porno atau dewasa. Dengan demikian, seandainya ada halaman berita pemerkosaan dan kesehatan yang sebenarnya tidak termasuk kategori porno pun dapat tetap terblokir.
3. Halaman website yang terblokir akan menampilkan halaman kosong dengan tulisan, “Maaf, halaman yang anda cari tidak dapat ditampilkan.”
4. Pengecekan hanya berdasarkan teks dalam halaman website yang dibuka. Image atau video tidak akan melewati pengecekan.

5. Pengecekan tidak dilakukan sampai level link berikutnya. Untuk dapat dilakukan pengecekan, maka link tersebut harus terbuka sampai muncul di url, baru pengecekan akan dilakukan.
6. Pengecekan dan pengkategorian website sebagai situs porno hanya dibatasi untuk website berbahasa Indonesia. Website berbahasa Inggris tetap akan mendapat hasil blokir atau tidak, namun keakuratan tidak terjamin.
7. Fitur dalam extension ini adalah adanya password. Bila password telah dipasang, maka untuk mengubah pengaturan atau *option* extension, pengguna harus memasukkan password. Fitur lainnya yang disediakan adalah fitur blokir website tertentu, dan memberi pengecualian terhadap website tertentu.
8. Aplikasi ini berbasis web dengan bahasa pemrograman Javascript untuk extension Google Chrome, PHP untuk modul *stemming*, dan Visual Basic dengan framework ASP.NET untuk web service.
9. Perangkat lunak yang digunakan dalam pembuatan sistem adalah Microsoft Visual Studio 2010. *Database* pada komputer *server* menggunakan Microsoft SQL Server 2008. *Web server* menggunakan WampServer.

1.4 Tujuan Penelitian

Tujuan dari pengenalan tanda tangan dengan pengolahan citra digital dan penggunaan jaringan saraf tiruan ini adalah:

1. Mencegah adanya konten porno atau dewasa untuk tidak dapat dibaca oleh anak-anak.
2. Mengetahui ketepatan klasifikasi berdasarkan pelatihan menggunakan metode Naive Bayes.
3. Menggabungkan beberapa metode seperti *stemming* Sastrawi dan algoritma klasifikasi Naive Bayes untuk dapat dimanfaatkan menjadi *content filtering*.

1.5 Manfaat Penelitian

Manfaat yang hendak dicapai dalam penelitian ini dibagi atas 2 yaitu:

1. Manfaat Teoritis

Penelitian ini diharapkan dapat menjadi referensi penelitian selanjutnya berkaitan dengan pemblokiran halaman website porno berdasarkan teks dan klasifikasi dengan Naive Bayes.

2. Manfaat Praktis

Hasil tugas akhir ini diharapkan dapat membantu orang tua dalam memastikan anak-anaknya tidak membuka website-website yang mengandung konten porno atau dewasa. Masih banyak konten porno atau dewasa yang luput dari pengawasan pemerintah, dan dibutuhkan filter selanjutnya yang dapat membantu orang tua membatasi tampilan internet untuk anak-anak. Oleh karena itu hasil dari tugas akhir ini diharapkan dapat memperkecil resiko anak-anak membaca konten yang tidak layak mereka lihat.

1.6 Sistematika Penulisan

Untuk mempermudah dalam memahami sistematika penulisan, penelitian ini dibagi menjadi beberapa bab, yaitu:

I. PENDAHULUAN

Bab ini berisikan latar belakang permasalahan yang akan dibahas dalam skripsi ini, kemudian berisi batasan masalah serta fokus penelitian yang akan diteliti, tujuan dan juga manfaat dari penelitian ini.

II. PEMBLOKIRAN WEBSITE

Bab ini menjelaskan teori-teori yang diperlukan dalam pemblokiran halaman konten porno dan dewasa, yaitu mengenai pornografi, *content filtering*, *browser extension*, *tokenization*, *stopword removal*, *stemming*, dan Naive Bayes. Pembahasan dalam bab ini akan disertai dengan berbagai literatur yang berkaitan dengan teorinya.

III. ALUR KLASIFIKASI WEBSITE

Bab ini menjelaskan mengenai rancang alur, metode pengambilan dan pengolahan teks halaman website serta pengklasifikasian dengan perhitungan Naive Bayes.

IV. IMPLEMENTASI PROGRAM

Bab ini membahas tentang implementasi dan pengujian dari hasil analisa dan perancangan yang telah disusun pada bab sebelumnya.

V. PENUTUP

Bab ini merupakan penutup penyusunan laporan yang berisi kesimpulan dari implementasi dan uji coba yang dilakukan. Selain itu berisi pula saran yang diharapkan dapat menjadi masukan untuk pengembangan aplikasi di masa yang akan datang.