PERANCANGAN KEAMANAN AKSES INTERNET BERBASIS TEXT FILTERING PADA UNIVERSITAS ATMA JAYA MAKASSAR

Harry Susanto Wirawan
Prodi Teknik Informatika, Fakultas Teknologi Informasi, Universitas Atma Jaya Makassar
Alamat e-mail: harryswirawan@gmail.com

ABSTRACT

University of Atma Jaya Makassar has good internet access that can be used by everyone in the environment. From the observations, it is known that networks that can be accessed openly have not yet been handled to filter content on a website. Starting from this, the researcher intends to design a proxy server that can handle filtering text content and also optimizes internet access security on UAJM. The results of this study showed that sites categorized as pornography and radicalism are taken at random and then evaluated giving result of 74.9% for pornographic websites and 20.1% for radicalism websites.

Keywords: *filtering, security, text content, text matching, squidanalyzer.*

1. PENDAHULUAN

Saat ini kebutuhan akan layanan Internet semakin meningkat. Hal ini ditandai dengan makin beragam usia para penggunanya, mulai dari golongan anak-anak hingga dewasa yang menggunakan internet guna mendapatkan informasi yang mereka inginkan. Internet pun menyediakan berbagai jenis layanan informasi, mulai dari berita, dunia hiburan, hingga ilmu pengetahuan yang disajikan dalam bentuk text, gambar, maupun video. Dari data yang di ambil dari Asosiasi Penyelenggara Jasa Internet Indonesia (APJII) pada tahun 2017 menunjukkan populasi penduduk di Indonesia ada 262 juta orang dan lebih dari setengahnya yaitu sekitar 143 juta orang telah terhubung dengan internet. Kemunculan internet yang begitu canggih membawa kemudahan kepada para pengguna untuk mengakses informasi yang dibutuhkan. Dari penggunaan internet yang sangat banyak tersebut salah satu masalah yang muncul adalah konten negatif yang tersebar luas di internet, berdasarkan data yang diambil dari Menkominfo hingga awal oktober 2018, sudah ada lebih dari 890.000 website yang berisikan konten negatif yang telah diblokir dan 80% diantaranya adalah website pornografi dan selama tahun 2019 Kemenkominfo telah memblokir sebanyak 1.500 situs yang berkonten radikalisme dan terorisme.

Berdasarkan Peraturan Menteri Komunikasi dan Informasi (Menkominfo) No. 19 Tahun 2014, konten negatif meliputi pornografi dan kegitan illegal lainnya berdasarkan ketentuan perundang-undangan. Khusus untuk pornografi pada UU No. 44 tahun 2008 telah mendifinisikan pornografi merupakan gambar sketsa, ilustrasi, foto, tulisan, suara, bunyi, gambar bergerak, animasi, kartun, percakapan, gerak tubuh, atau bentuk pesan lainnya melalui berbagai bentuk media komunikasi dan/atau pertunjukan dimuka umum, yang membuat kecabulan atau eksploitasi seksual yang melanggar norma kesusilaan dalam masyarakat. Kemudian untuk radikalisme, dikutip dari Menteri Koordinator bidang politik, hukum, dan keamanan, Bapak Mahfud MD menegaskan, definisi radikalisme yang digunakan pemerintah merujuk pada Undang-Undang Nomor 5 tahun 2018 tentang Pemberantasan Tindak Pidana Terorisme, yang dimana disebutkan bahwa radikalisme itu tindakan melawan hukum untuk mengubah sistem, bukan secara gradual melainkan secara radikal, dengan cara kekerasan. Salah satu contoh dampak dari radikalisme bagi perguruan tinggi yaitu peristiwa pengeboman bunuh diri di Surabaya pada tanggal 13-14 Mei 2018, yang dimana meluasnya pembicaraan di kalangan publik tentang meningkatnya paham radikal di Perguruan Tinggi Negeri (PTN). Pembicaraan dan perdebatan ini berawal dari

adanya pernyataan dari beberapa dosen dan termasuk diantaranya profesor yang seolah-olah merestui aksi bom bunuh diri tersebut. Selain itu ada juga dosen dan profesor PTN yang mendukung pemahaman dan praksis yang ingin dibentuk dakwah Islamiyah atau khilafah, pemahaman dan pemikiran ini baik secara langsung maupun tidak langsung menolak NKRI dan Pancasila.

Universitas Atma Jaya Makassar sendiri (UAJM) saat ini telah memiliki layanan koneksi internet yang diperuntukkan bagi para karyawan, dosen, dan mahasiswa, baik melalui teknologi Wi-Fi maupun koneksi via kabel LAN. Dari hasil observasi yang dilakukan di UAJM pada tahun 2018, ditemukan beberapa kekurangan terhadap pengamanan akses internet terhadap website-website yang disinyalir memuat konten negatif yang dalam penelitian ini mencangkup konten pornografi dan radikalisme. Berdasarkan daftar website yang terblokir dari Kominfo pada tahun 2018, ada sembilan ratus tiga puluh tiga ribu enam ratus tujuh puluh satu website negatif yang telah di blokir dan 9.4% di antaranya hanya menggunakan alamat IP (*Internet Protocol*) saja, website tersebut sulit di tangani karena tidak menggunakan DNS (*Domain Name System*).

Dari Masalah yang di paparkan di atas, maka perancangan keamanan akses internet berbasis konten text filter pada Universitas Atma Jaya Makassar dibutuhkan, karena perancangan ini sesuai dengan misi yang dikutip dari website UAJM yaitu "Mengelola pendidikan tinggi dalam suasana akademik yang beretika dan bermartabat". Dari tujuan perancangan tersebut penelitian ini akan menggunakan proxy, karena proxy sendiri adalah salah satu cara untuk melakukan filtering pada jaringan internet dengan memblokir isi konten, url dari suatu website, file, dan juga kata kunci dengan menggunakan web scraping yang dipadukan dengan algoritma text filtering, metode classification dan juga pembobotan kata karena akan lebih optimal jika di terapkan di UAJM saat ini. Dengan diterapkannya perancangan ini diperkirakan dapat membantu penggunaan internet sesuai kebutuhan setiap bagian di Universitas Atma Jaya Makassar.

Saat ini sudah ada penelitian yang mengangkat perancangan untuk akses internet yaitu penelitian Asfato (2017), dan Ronny (2017). Berdasarkan dari data yang diperoleh dari penelitian sebelumnya pada saat itu jumlah mahasiswa aktif ada sekitar 1.400 dan jumlah dosen atau karyawan sebanyak 200 orang. Pada saat penelitian tersebut dibuat total bandwith di UAJM adalah 10MB/s dengan 7MB/s untuk traffic nasional (eXchange (IIX)) dan 3 MB/s untuk traffic Internasional (eXchange (IX)) berlaku dari September 2016 dan penggunaan bandwith download perhari adalah 679.855 KiB dan kecepatan rata-rata download sebesar 142.372 KiB. Penelitian sebelumnya telah berhasil menghasilkan sebuah sistem manajemen akses jaringan berbasis bandwith dan user. Namun penelitian tersebut masih mengandalkan filter website dari sisi Internet Service Provider (ISP) yaitu dengan memblok (domain name system) DNS yang sifatnya statis dan belum menggunakan filter cerdas untuk mengenali isi konten teks dalam suatu website yang saat ini sangat penting karena sudah banyak website dengan isi konten negatif yang pada penelitian ini yaitu situs pornografi dan radikalisme yang tidak menggunakan DNS atau hanya menggunakan ip address saja sebagai alamat websitenya.

Penelitian ini sangat penting dilakukan agar mahasiswa,dosen ataupun staff yang berada pada lingkungan UAJM dapat menjalankan proses perkuliahan ataupun pekerjaan dengan suasana yang beretika dan bermatabat yaitu dengan menjauhkan pengguna jaringan di lingkungan UAJM dari website – website pornografi dan radikalisme yang dimana hal tersebut bisa berdampak buruk, diantaranya jika pengguna menganut faham radikalisme itu dapat meresahkan banyak orang, menghilangkan keharmonisan antar umat beragama, mencoreng nama baik suatu agama, dan dampak dari pornografi diantaranya orang dapat lebih mudah berbohong, pendidikan yang terganggu, terjadinya penyimpangan seksual, dan mudah depresi dan cemas.

2. TINJAUAN PUSTAKA

2.1 *Proxy Server*

Proxy dalam pengertiannya sebagai perantara, bekerja dalam berbagai jenis protokol komunikasi jaringan dan dapat berada pada level-level yang berbeda pada hirarki layer protokol komunikasi jaringan.

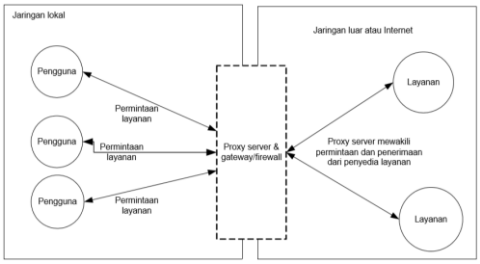
Suatu perantara dapat saja bekerja pada layer Data-Link, layer Network dan Transport, maupun layer Aplikasi dalam hirarki layer komunikasi jaringan menurut OSI. Namun pengertian Proxy Server sebagian besar adalah untuk menunjuk suatu server yang bekerja sebagai Proxy pada layer Aplikasi(Arjuni, 2010) [1].

Ada 2 jenis proxy yang paling sering digunakan mempunyai fungsi masing-masing, dan tiap proxy tersebut digunakan dengan tujuan yang berbeda-beda:

1. Proxy Transparan (*Transparent Proxy*)
   Alamat IP client dapat terdeteksi oleh server tujuan (server provider). Proxy jenis ini sangat sering digunakan untuk meningkatkan kecepatan Internet.
2. Proxy Anonim (Anonymous Proxy)
   Alamat IP client tidak terdeteksi oleh server provider Internet, namun provider mengetahui apabila koneksi dilakukan melalui proxy. Proxy jenis ini sangat berguna sekali saat digunakan untuk menjaga privasi IP address clients saat melakukan *browsing*.

2.2 Komunikasi Data Berbasis Client-Server

Prinsip kerja dalam komunikasi data pada proxy server ialah, saat user menggunakan layanan suatu proxy kemudian meminta file atau data yang terdapat di public server maka proxy akan meneruskannya ke internet jadi seolah-olah proxy tersebut yang memintanya. Dan saat proxy server telah mendapatkan apa yang diminta oleh user, proxy akan memberikan respon kepada user jadi seolah-olah dialah publik servernya (firmansyah & riadi, 2014) [2].



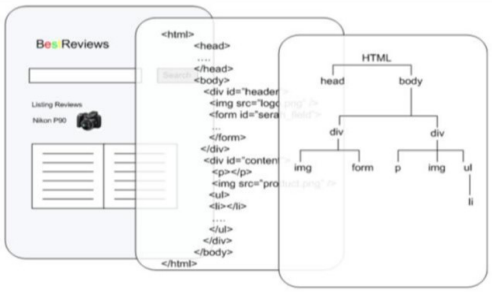
Gambar 1. Komunikasi Data Proxy Server

2.3 *Web Scraping*

Menurut Saurkar, Pathare, dan Gode (IJFRCSCE, 2018:364) [3] web scraping adalah teknik penting yang digunakan untuk mengekstrak data tidak terstruktur dari situs web dan mengubah data menjadi terstruktur. Web Scraping juga diidentifikasi sebagai web data extraction, web data scraping, web harvesting atau screen scraping. Web scraping adalah bentuk data mining. Tujuan dasar dan penting dari proses web scraping adalah untuk mengambil informasi dari situs web yang berbeda dan tidak terstruktur dan mengubahnya menjadi informasi struktur yang dapat dipahami seperti spreadsheet, database, atau file CSV. Data seperti harga barang, harga saham, laporan berbeda, harga pasar dan detail produk, dapat dikumpulkan melalui web scraping.

Gambar 2. Arsitektur Dasar Web Scraping

Scraping adalah teknik yang digunakan untuk memotong informasi dari halaman web berdasarkan rutinitas skrip. Halaman web adalah dokumen yang ditulis dalam Hypertext Markup Language (HTML), dan yang terbaru adalah XHTML yang berbasis XML. Dokumen web diwakili oleh tree structure yang disebut Document Object Model (DOM), atau hanya DOM tree dan tujuan HTML adalah untuk menentukan format teks yang ditampilkan oleh browser web seperti yang ditunjukkan pada gambar berikut.

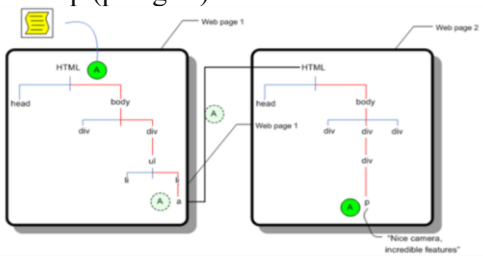


Gambar 3. Dokumen Web, kode HTML dan Document Object Model

Dari sudut pandang operasi, web scraping terlihat seperti salin dan kemudian tempel secara manual. Perbedaannya adalah pekerjaan ini dilakukan secara terorganisir dan otomatis berkat peran virtual computer agent. Ketika seorang agen mengikuti setiap tautan halaman web, ia sebenarnya melakukan operasi yang sama seperti yang biasa dilakukan manusia ketika berinteraksi dengan situs web. Agen ini dapat mengikuti

tautan (dengan mengeluarkan permintaan GET HTTP) dan mengirimkan formulir (melalui HTTP POST), menjelajah banyak halaman web yang berbeda. Sekarang manfaatnya tampak jelas ketika pengguna harus mengklik beberapa tautan sebelum sampai ke halaman yang diinginkan. Langkah selanjutnya adalah, parser mengikuti jalur yang ditentukan pengguna di dalam dokumen untuk mengambil informasi yang diinginkan berdasarkan data yang diambil pada langkah sebelumnya. Jalur ini ditentukan oleh CSS selectors atau XPATH.

Mereka menggunakan path relatif dan absolut (berdasarkan DOM tree) untuk mengarahkan parser ke elemen tertentu di dalam dokumen web. Biasanya operasi web scraping menggunakan ekspresi reguler untuk mempersempit atau memangkas informasi yang ditemukan, untuk mengambil data yang ditentukan pengguna. Proses ini diilustrasikan dalam Gambar 4. Web scraping agent mengumpulkan informasi dari halaman web - lingkaran bertitik mewakili web scraping agent yang melintasi DOM tree. Garis merah adalah XPATH untuk elemen yang diinginkan dalam dokumen. Agen mencapai hyperlink di halaman web 1 dan melanjutkan ke halaman web 2 hingga menemukan informasi yang dilampirkan oleh elemen p (paragraf).



Gambar 4. Langkah Mengumpulkan Informasi

2.4 Penyaringan Konten

Dikutip dari Goset dan Shorter (2011) [4], Penyaringan konten (juga dikenal sebagai penyaringan informasi) adalah penggunaan program untuk menyaring dan mengecualikan dari akses atau halaman web ketersediaan atau email yang dianggap tidak pantas. Pemfilteran konten digunakan oleh perusahaan sebagai bagian dari komputer firewall Internet dan juga oleh pemilik komputer rumah. Pemfilteran konten biasanya bekerja dengan menentukan string karakter, jika dicocokkan, menunjukkan konten yang tidak diinginkan yang akan disaring. Konten biasanya disaring untuk konten pornografi dan terkadang juga untuk konten yang berorientasi pada kekerasan atau kebencian. Kritik terhadap program penyaringan konten menunjukkan bahwa tidak sulit untuk tidak sengaja mengecualikan konten yang diinginkan.

2.5 Dasar Teori Pornografi dan Radikalisme

Pada UU No. 44 tahun 2008 telah mendifinisikan bahwa Pornografi merupakan gambar sketsa, ilustrasi, foto, tulisan, suara, bunyi, gambar bergerak, animasi, kartun, percakapan, gerak tubuh, atau bentuk pesan lainnya melalui berbagai bentuk media komunikasi dan/atau pertunjukan dimuka umum, yang membuat kecabulan atau eksploitasi seksual yang melanggar norma kesusilaan dalam masyarakat.

Sedangkan radikalisme berasal dari bahasa Latin radix yang berarti akar. Yaitu berpikir secara mendalam terhadap sesuatu sampai ke akar-akarnya. Radikalisme merupakan suatu paham yang menghendaki adanya perubahan, pergantian, dan penjebolan terhadap suatu sistem masyarakat sampai ke akarnya. Radikalisme menginginkan adanya perubahan secara total terhadap suatu kondisi atau semua aspek kehidupan masyarakat. Kaum radikal menganggap bahwa rencana-rencana yang digunakan adalah rencana yang paling ideal. Terkait dengan radikalisme ini, seringkali beralaskan pemahaman sempit agama yang berujung pada aksi terror bom (Rohimah, 2017) [5].

2.6 Daftar Kata Pornografi dan Radikalisme

Daftar kata yang terkumpul sebagai daftar kata pornografi dan radikalisme adalah sebanyak 381 kata untuk pornografi dan 171 kata untuk radikalisme. Kata-kata yang dikategorikan sebagai kata radikalisme diambil dari penelitian yang telah dilakukan sebelumnya oleh M. Subhan, A.Sudarsono, A.Barakbah, 2017, "*Classification of Radical Web Content in Indonesia using Web Content Mining and k-Nearest Neighbor Algorithm*" [6] dan kata yang dikategorikan sebagai kata pornografi di ambil dari penelitian sebelumnya yang dilakukan oleh Lubis,

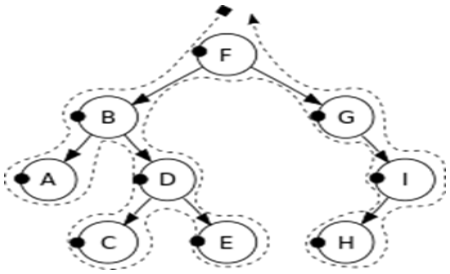
S.Husen, 2018, "Membangun Korpus Pornografi Bahasa Indonesia" [7]. Kata-kata yang diambil terdiri dari bahasa indonesia dan inggris. Kata-kata tersebut yang dijadikan referensi daftar kata dalam proses text matching nantinya.

2.7 Hubungan *Content Filtering* dan *Web Scraping*

Berdasarkan konsep yang ingin dibuat oleh penulis, kaitan antara content filtering dengan *web scrapin*g yaitu dengan memanfaatkan log proxy server sebagai media untuk mengambil isi url yang diakses oleh pengguna jaringan, lalu data url tersebut akan di ekstrak menggunakan web scraping. Kemudian data yang sudah di ekstrak dan sudah menjadi data terstruktur akan di filter dan di bandingkan dengan data kata negatif yang telah di input admin. Setelah proses perbandingan maka tinggat kecocokan tersebut akan dikalkulasikan menggunakan proses *classification* untuk menentukan apakah website tersebut bisa di golongkan sebagai website berisi konten negatif. Adapun penanganan untuk scraping website yang menggunakan HTTPS peneliti menggunakan salah satu library dari bahasa pemrograman python, yaitu BeautifulSoup dalam membuka url tersebut sebelum di scraping.

2.8 Algoritma *Recursive Backtracking*

Algoritma *backtracking*, algoritma yang berbasis pada DFS (*Deep-First Search*) untuk mencari solusi persoalan secara lebih akurat (Sulistyowati, 2017). Algoritma ini mengkhususkan diri dalam membangun solusi satu per satu secara bertahap sambil menghilangkan solusi yang gagal memenuhi kendala. Dengan algoritma backtracking kita tidak perlu memeriksa semua kemungkinan solusi yang ada, hanya pencarian yang mengarah ke solusi saja yang selalu dipertimbangkan. Akibatnya, waktu pencarian dapat dihemat. Contoh pencarian solusi yang dilakukan dengan cara menelusuri node-node dari sebuah tree secara pre-order.



Gambar 5. Pencarian Solusi Algoritma *Backtracking*

Algoritma *backtracking* mempunyai prinsip dasar yang sama seperti brute-force yaitu mencoba segala kemungkinan solusi. Perbedaanya ialah semua solusi dibuat dalam bentuk pohon solusi dan algoritma akan menelusuri pohon tersebut secara DFS sampai menemukan solusi yang layak.

2.9 Kajian Penelitian Sejenis

Penelitian mengenai text filtering telah banyak dilakukan. Penelitian tersebut dilakukan dengan metode berbeda, seperti menggunakan web scraping, menggunakan web data mining, ataupun OCR (*Optical Character Recognition*). Berikut penelitian sejenis yang pernah dilakukan :

1. Ma'arif, Muhammad Rifqi (2016) dengan judul "Integrasi Laman Web Tentang Pariwisata Daerah Istimewa Yogyakarta Memanfaatkan Teknologi Web Scraping dan Text Mining" [8]. Penelitian ini bertujuan untuk membangun sebuah laman web yang mampu mengintegrasikan informasi dari laman-laman web yang lain yang memuat informasi mengenai pariwisata DIY. Integrasi informasi akan dibuat dengan memanfaatkan teknologi web scraping dan text mining. Dengan adanya laman web yang mengintegrasikan informasi dari laman-laman web yang lain, calon wisatawan tidak perlu lagi menghabiskan banyak waktu untuk mencari informasi pariwsata DIY yang lengkap dan akurat, dengan mengekspolarasi laman-laman yang menyajikan informasi mengenai pariwisata DIY kemudian mengumpulkan semua informasi yang ada di laman tersebut secara otomatis dengan menggunakan teknologi web scraping. Informasi yang sudah terkumpul kemudian akan diolah dengan menggunakan teknologi text mining

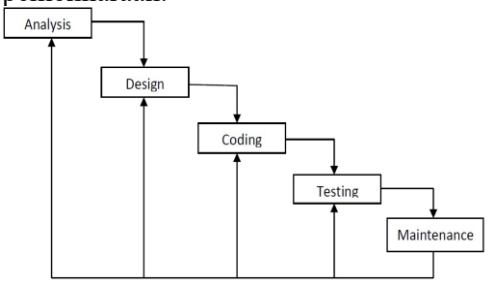
sehingga menjadi informasi yang lebih ringkas namun lengkap dan akurat untuk disajikan kepada para calon wisatawan.

2. Vargiu (2013) dengan judul "Exploiting web scraping in a collaborative filtering - based approach to web advertising" [9]. Penelitian ini berfokus pada teknik untuk mengekstraksi konten halaman web yaitu dengan mengadopsi teknik memo di bidang periklanan web dengan cara menemukan iklan yang paling relevan untuk halaman web umum dengan mengeksploitasi web scraping.

3. METODOLOGI PENELITIAN

Berdasarkan penelitian yang dilakukan peneliti, maka penelitian ini merupakan jenis penelitian terapan, dimana penelitian ini didasarkan dari kenyataan-kenyataan praktis, penerapan, dan pengembangan ilmu pengetahuan yang dihasilkan dari penelitian dasar dalam kehidupan nyata, bertujuan agar dapat melakukan sesuatu yang lebih baik.

Metode perancangan yang digunakan pada penelitian ini menggunakan metode Waterfall yang dimulai dari tahap analisis, perancangan, pembuatan, pengujian, dan pemeliharaan.



Gambar 6. Metode Waterfall

Tahapan – tahapan dalam metode Waterfall sebagai berikut:

1. *Analysis*
   Pada tahap ini akan dilakukan proses analisa terkait prasyarat minimum dari sebuah content filtering berbasis teks untuk dapat bekerja. Untuk itu, diperlukan data seperti bagaimana topologi jaringan yang diterapkan sekarang di UAJM, kemudian melakukan proses wawancara kepada BAPSI sebagai pengelola layanan jaringan pada UAJM agar peneliti dapat mengetahui hal-hal apa saja yang diperlukan dalam merancang sistem keamanan akses internet.
2. *Design*
   Tahap ini merupakan proses yang akan berfokus pada desain pembuatan konten filtering berbasis teks dengan design awal yaitu melakukan penginstalan proxy server, konfigurasi jaringan, penerapan text filtering pada proxy server dan penerapan ACL untuk url. Rancangan desain dibuat menggunakan diagram konteks, diagram berjenjang, dan data flow diagram (DFD) pada sistem keamanan akses internet.
3. *Coding*
   Pada tahap ini akan dilakukan pengembangan dan penerapan beberapa algoritma dalam melakukan text filtering dengan menerapkan metode classification pada url yang diakses oleh pengguna jaringan di UAJM dengan menggunakan bahasa program python untuk diterapkan pada sistem keamanan akses internet.
4. *Testing*
   Pada tahap ini dilakukan pengujian untuk menemukan kesalahan-kesalahan yang mungkin terjadi pada saat proses filtering berjalan dan memastikan juga bahwa algoritma yang diterapkan untuk melakukan text filtering dapat berjalan sebagaimana mestinya. Pengujian sistem akan dilakukan dengan metode whitebox.
5. *Maintenance*
   Tahap ini dilakukan setelah perancangan telah digunakan oleh pengguna. Perubahan dilakukan jika terdapat kesalahan, oleh karena itu perancangan akan disesuaikan lagi untuk menyesuaikan perubahan kebutuhan yang diinginkan pengguna.

4. HASIL DAN PEMBAHASAN

4.1 Penentuan Kebutuhan

Dalam menentukan kebutuhan terdapat beberapa hal yang harus ditempuh agar dapat menerapkan keamanan akses internet dengan metode filtering text pada pengguna jaringan dalam lingkup UAJM. Untuk merancang keamanan akses internet dengan metode filtering text yang tepat dan juga sesuai dengan kebutuhan di UAJM, maka peneliti melakukan metode wawancara untuk menentukan kebutuhan apa saja yang

diperlukan dalam penerapan penelitian ini di UAJM.

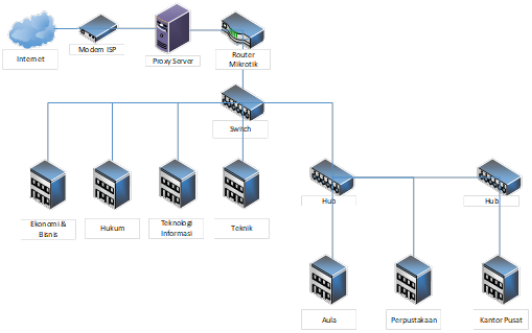
Dari hasil wawancara didapatkan bahwa jaringan yang sudah ada masih kurang stabil jika diakses secara bersamaan oleh mahasiswa, staff, dan juga dosen, dikarenakan arsitektur yang masih kurang memadai untuk saat ini. Kemudian untuk segi keamanan sudah diterapkan sistem login terlebih dahulu sebelum pengguna dapat mengakses jaringan internet, tetapi belum ada fitur untuk menyeleksi kata kunci, teks ataupun url yang mengandung kata negatif seperti pornografi, sara, ujaran kebencian dan lainnya selain dari fitur filter yang diberikan oleh ISP (Internet Service Provider). Maka dari itu peneliti ingin menerapkan keamanan filtering text sangat membantu untuk mengatasi kekurangan dari segi keamanan yang sudah ada.

4.2 Desain

4.2.1 Rancangan Topologi Jaringan

Dalam penerapan text filtering dalam pengamanan akses internet di UAJM sekarang sama sekali tidak mengubah topologi yang sudah diterapkan, penerapannya hanya menambah satu server proxy eksternal yang terhubung ke router utama, tujuannya agar semua traffic dari pengguna yang terkoneksi bisa di monitoring dan teks dari website yang di akses bisa di filter dengan baik.

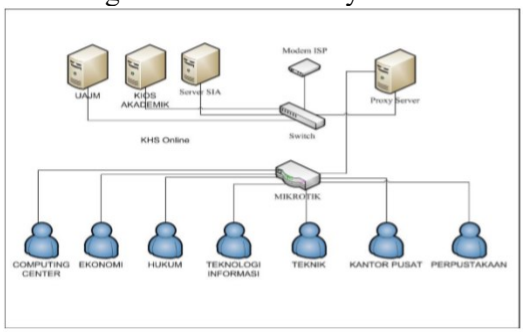
Penggunaan proxy server dalam penerapan text filtering sangat dibutuhkan untuk membantu memberikan log informasi seperti website apa saja yang di akses oleh pengguna yang terkoneksi pada jaringan di UAJM.



Gambar 7. Topologi Internet Setelah Menggunakan Proxy Server

4.2.2 Rencana Integrasi

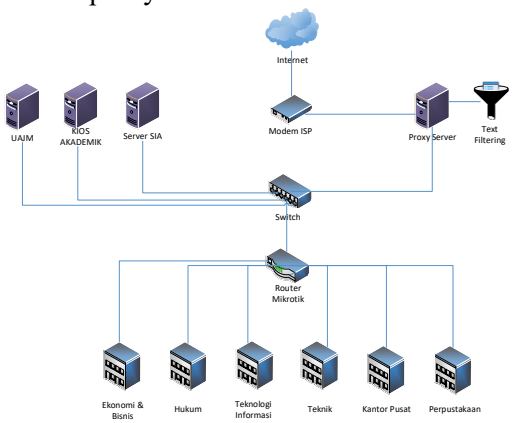
Berdasarkan konsep perencanaan dalam mengintegrasikan sistem yang telah dirancang oleh peneliti yaitu "Perancangan Sistem Keamanan Akses Internet Berbasis Text Filtering pada Universitas Atma Jaya Makassar" dengan sistem yang telah dibuat sebelumnya oleh saudara Ronny Theodorus yaitu "Manajemen Bandwith dan Pengguna Jaringan pada Universitas Atma Jaya Makassar" dan sistem yang dibuat oleh saudara Asfato Loe yaitu " Perancangan Server Proxy pada Universitas Atma Jaya Makassar", ialah dengan menambahkan sistem yang telah dibuat oleh peneliti ke sistem yang telah dibuat oleh saudara Asfato Loe dan saudara Ronny Theodorus sebelumnya, yang dimana sistem yang telah dirancang oleh saudara Asfato menggunakan Proxy Server dan sistem yang telah dirancang oleh saudara Ronny menggunakan Mikrotik dalam penerapannya. Proxy server sendiri digunakan oleh peneliti dalam merancang sistem text filtering ini, yang dimana sistem yang telah dibuat oleh peneliti, proxy server berguna dalam mengambil URL yang diakses oleh pengguna yang terkoneksi pada jaringan di UAJM. Gambaran rancangan sebelumnya dari saudara Asfato dapat dilihat pada gambar 8, yang dimana di proxy Server terdapat sistem yang dirancang oleh saudara Asfato, dan di mikrotik terdapat sistem yang telah dirancang oleh saudara Ronny.



Gambar 8. Desain Integrasi Sistem Sebelumnya

Pada gambar 9 dapat dilihat rancangan yang telah dibuat oleh peneliti yaitu sistem text filtering akan melengkapi sistem yang telah dirancang oleh saudara Asfato yang terdapat pada proxy server dengan menambahkan fungsi pemfilteran teks agar penanganan atas website – website negatif dengan kategori pornografi dan radikalisme bisa lebih optimal. Dalam rancangan ini, koneksi dari modem ISP akan langsung masuk ke proxy server dan berada paling

depan yang seolah – olah bertindak sebagai komputer lainnya untuk melakukan *request* terhadap konten dari internet, yang dimana rancangan sebelumnya request keluar dan masuk yang dilakukan tidak semuanya melalui proxy server terlebih dahulu.



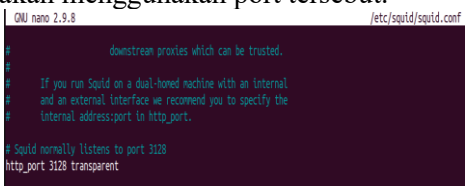
Gambar 9. Desain Integrasi Sistem

4.3 Pengembangan Sistem

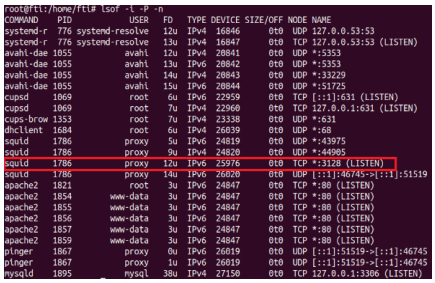
4.3.1 Proses Konfigurasi awal *Proxy Server*

Berdasarkan dari hasil desain jaringan yang telah dibuat oleh peneliti, maka langkah selanjutnya yang akan dilakukan ialah menginstall dan mengkonfigurasi proxy server. Konfigurasi proxy server dilakukan agar dapat mendukung dalam proses penerapan sistem keamanan akses internet berbasis text filtering yang dimana proxy server menjadi penerima informasi dari situs apa saja yang diakses oleh pengguna jaringan pada UAJM.

1. Konfigurasi Port Proxy Server
   Proses konfigurasi port pada gambar 10, adalah adalah tahap awal untuk menetapkan port yang nantinya akan menjadi listener dalam melakukan koneksi dalam jaringan. Kemudian pada gambar 11 dapat dilihat setelah konfigurasi port proxy sudah berjalan dalam mode listen, yang artinya port tersebut menunggu sebuah koneksi yang akan menggunakan port tersebut.

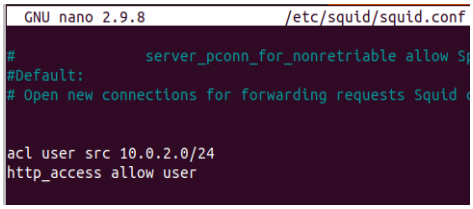


Gambar 10. Konfigurasi Port Proxy Server



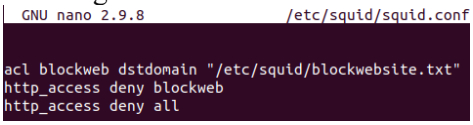
Gambar 11. Port Proxy Server yang Berjalan

2. Konfigurasi IP
   Proses konfigurasi IP ini adalah tahap awal untuk menetapkan IP yang digunakan menjadi IP *source* yang akan menjadi perantara dari semua *traffic* jaringan yang berasal dari pengguna jaringan pada UAJM sebelum diteruskan ke internet.



Gambar 12. Konfigurasi IP Proxy Server

3. Konfigurasi ACL (Access Control List)
   Tahap berikutnya yaitu membuat ACL untuk situs yang akan terblokir nantinya, disini peneliti membuat satu file *text* yang yang akan berisi url yang akan dikirim dari sistem, yang dimana url tersebut sudah diklasifikasikan sebagai situs negatif.



Gambar 13. Konfigurasi ACL

4.3.2 Proses *Scraping Text*

Pada proses ini peneliti melakukan pembuatan program python yang menangani proses scraping text yang urlnya diperoleh dari log proxy yang berisikan data url yang telah diakses oleh pengguna jaringan.

1. Proses ekstrak URL dari log proxy
   Proses ini adalah tahap pertama dalam *scraping text* yaitu dengan mengekstrak url valid yang terdapat pada log proxy lalu menyeleksi url tersebut, kemudian daftar url tersebut akan disimpan ke file teks yang nantinya akan di akses pada proses selanjutnya.

```
def write():
        with open("/var/log/squid/access.log") as file:
                for string in file:
                        extractor = URLExtract()
                        urls = extractor.find_urls(string)
                        sys.stdout = open('web1.txt','a')
                        print (urls)
                        sys.stdout.close()
                        fin = open("web1.txt", "rt")
                        data = fin.read()
                        data = data.replace('http://', '')
                        data = data.replace(':443', '')
                        data = data.replace("['", "")
                        data = data.replace("']", "")
                        data = data.replace("\00","")
                        fin = open("web1.txt", "wt")
                        fin.write(data)
                        fin.close()
                        lines = open('web1.txt', 'r').readlines()
                        lines_set = set(lines)
                        out  = open('web1.txt', 'w')
                        for line in lines_set:
                                out.write(line)
```

Gambar 14. Proses Ekstrak URL dari log proxy

2. Proses Ekstrak teks

   Proses selanjutnya yaitu mengekstrak teks yang terdapat dari website yang akan diakses dari file yang berisi daftar url yang telah diperoleh dari log proxy. Tahap pertama yaitu membuka file yang berisi daftar url, lalu url tersebut akan akses oleh sistem menggunakan protokol http atau https, setelah url terbuka website tersebut akan di scraping menggunakan *library* yang terdapat di kode program python yaitu beautifulsoup, disini peneliti menggunakan beautifulsoup4. Setelah sistem menerima teks yang terdapat di website yang diakses, sistem akan menyeleksi teks yang dianggap valid. Setelah proses seleksi teks tersebut yang terakhir dilakukan yaitu menggabungkan teks tersebut dan kemudian akan disimpan sementara kedalam salah satu type data yang terdapat di python yaitu lists, yang nantinya akan dibandingkan dengan *data training* untuk mendapatkan hasil klasifikasi seperti website tesebut akan diblokir atau diperbolehkan untuk diakses oleh pengguna jaringan.

```
def scrap():
        line = open("web1.txt" ,"r")

        for url in line:
                        try :
                                r  = urlopen("http://" + url, timeout=5)
                except (Exception, urllib.error.URLError,urllib.error.HTTPError, socket.timeout):
                                continue
                try :
                                r = urlopen("https://" + url,timeout=5)
                except (urllib.error.URLError,urllib.error.HTTPError, socket.timeout):
                                pass
                except (ssl.CertificateError):
                                continue
                soup = BeautifulSoup(r,"lxml")
                for script in soup (["script", "style"]):
                                script.extract()

                text = soup.get_text()
                lines = (line.strip() for line in text.splitlines())
                chunks = (phrase.strip() for line in lines for phrase in line.split("  "))
                text = re.sub(r'([a-z](?=[A-Z])|[A-Z](?=[A-Z][a-z]))', r'\1 ', text)
                text = re.sub(r'[^A-Za-z0-9]+',' ', text)
                text = re.sub(' \d+', ' ', text)
                text = re.sub(r'\b\w{1,2}\b', '', text)
                text = text.replace('  ',' ')
                text = '\n '.join([ll.rstrip() for ll in text.splitlines() if ll.strip()])
                the_string = text
                kata = the_string.split()
                kata = [x.upper() for x in (kata)]
```

Gambar 15. Proses Ekstrak Teks

4.3.3 Proses Klasifikasi

Pada Proses ini peneliti membuat program python yang menangani proses klasifikasi dari situs yang diakses oleh pengguna yang sebelumnya telah diproses, kemudian mengkategorikannya sebagai situs yang aman untuk diakses ataupun tidak.

1. Proses Perbandingan dengan data *training*

   Pada proses ini daftar kata yang sudah di ekstrak dan di konversi ke tipe data *lists* pada proses sebelumnya, akan dibandingkan dengan data training yang berisi daftar kata negatif atau kata yang tidak pantas. Setelah proses perbandingan kedua daftar kata tersebut, yang selanjutnya dilakukan ialah menghitung berapa jumlah kata yang sama antara daftar kata yang telah di *scraping* dan *data training*. Proses perbandingan kata ini menggunakan *build-in package* dari python yaitu "regex" yang didalamnya terdapat fungsi *match* atau pencocokan kata yang menggunakan algoritma *recursive backtracking*.
2. Proses pengumpulan nilai bobot

   Proses berikutnya yang dilakukan ialah menghitung bobot tiap kata yang sama dengan data training pada proses sebelumnya, setiap kata yang sama tersebut kemudian diambil nilai bobotnya, yaitu 0.5 untuk bobot rendah, 1 untuk bobot sedang, dan 1.5 untuk bobot tinggi kemudian disimpan sementara di array untuk selanjutnya akan diproses kembali. Semakin tinggi nilai bobot dari suatu kata, maka semakin relevan kata tersebut muncul di website pornografi ataupun website radikalisme.
3. Proses pengumpulan kategori website

   Pada proses ini yang dilakukan adalah menghitung berapa kata pembanding di database yang sama dengan kata yang terdapat di sebuah website, yang dimana kata tersebut akan diambil kategorinya yaitu pornografi atau radikalisme.
4. Proses perhitungan kecocokan kata

   Proses yang akan dilakukan berikutnya yaitu mengkalkulasi tingkat kecocokan kata dari dua daftar kata yang dibandingkan sebelumnya. Proses kalkulasi ini menghitung nilai persentasenya yaitu dengan cara:

   ((nilai bobot rendah x jumlah kata bobot rendah) + (nilai bobot sedang x jumlah kata bobot sedang) + (nilai bobot tinggi

x jumlah kata bobot tinggi)) / (total kata / 2) x 100

Setelah mendapat nilai pesentase kecocokan kata antara daftar kata dari situs yang telah discrap dengan data training, maka nilai tersebut akan disimpan untuk diproses lagi pada fungsi selanjutnya.

5. Proses penentuan kategori website

   Proses penentuan kategori ini dilakukan dengan cara menghitung jumlah kategori dari kata yang sama dengan data training. Jika jumlah kata dengan bobot tinggi dijumlahkan dengan bobot sedang lebih besar dari jumlah kata kategori pornografi dan nilai kecocokan kata lebih besar dari 30% maka website tersebut akan dinyatakan sebagai website dengan kategori pornografi. Kemudian jika jumlah kata dengan bobot tinggi dijumlahkan dengan bobot sedang lebih besar dari jumlah kata kategori radikalisme dan nilai kecocokan kata lebih besar dari 10% maka website tersebut akan dinyatakan sebagai website dengan kategori radikalisme.

6. Proses pemblokiran dan *update* ACL proxy

   Proses pemblokiran ini dilakukan dengan mengecek kategori dari website yang telah diproses, jika kategori website tersebut tidak termasuk maka website tersebut aman dan tidak diblokir dan dapat diakses oleh pengguna yang terkoneksi ke jaringan UAJM. Tetapi jika selain dari kategori tidak termasuk, maka website tersebut akan diblokir dan url dari website tersebut akan *diupdate* ke ACL (Access Control List) proxy server agar pengguna jaringan yang terkoneksi tidak dapat mengakses situs tersebut.

4.3.4 Proses Evaluasi

Evaluasi dilakukan pada tanggal 2 Juli 2020 dengan menggunakan 10 website pornografi dan 10 website radikalisme yang diambil secara acak. Setelah website tersebut tidak terblokir, kemudian dihitung jumlah teks yang terdapat pada website yang telah dipilih. Berdasarkan hasil evaluasi yang dilakukan, hasil akurasi ialah sebesar 74.9% untuk 10 website pornografi dan 20.1% untuk 10 website radikalisme. Rincian evaluasi dari setiap website pornografi yang digunakan dapat dilihat pada tabel 1, sedangkan untuk website radikalisme dapat dilihat pada tabel 2.

Tabel 1. Evaluasi 10 Website Pornografi

No	Website	Persentase Kecocokan
1	192.243.98.23	65%
2	164.68.111.161	70%
3	167.99.74.239	100%
4	188.166.196.86	88%
5	www.pornhub.com	49%
6	www.forhertube.com	100%
7	www.xnxx.com	100%
8	www.redtube.com	61%
9	www.xvideos.com	65%
10	happy-porn.com	51%
Rata – rata persentase		**74.9%**

Tabel 2. Evaluasi 10 Website Radikalisme

No	Website	Persentase Kecocokan
1	al-khattab1.blogspot.com	31%
2	religionofallah.wordpress.com	14%
3	fadliistiqomah.blogspot.com	16%
4	jalanallah.wordpress.com	22%
5	daulah4islam.wordpress.com	13%
6	abdulloh7.wordpress.com	28%
7	mabesdim.wordpress.com	22%
8	kupastajam.blogspot.com	16%
9	jihadsabiluna-dakwah.blogspot.com	16%
10	anshardaulahislamiyahnusantara.wordpress.com	22%
Rata-rata persentase		**20.1%**

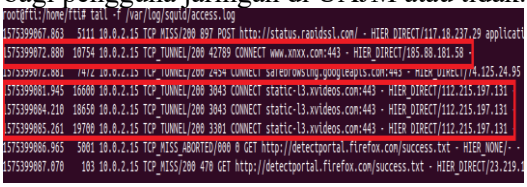
Untuk website radikalisme, batas persentase untuk dinyatakan terblokir yaitu 10%, ini dikarenakan website-website radikalisme memiliki banyak kata yang sulit untuk dikategorikan sebagai kata yang tergolong radikalisme dan juga dari hasil evaluasi memperlihatkan rata-rata persentase 20.1% dan persentase perbandingan website radikalisme yang terendah yaitu 13%, maka dari itu peneliti menetapkan batas persentase untuk website radikalisme diblokir yaitu 10%.

4.4 Uji Simulasi Sistem

Keamanan dalam mengakses internet di UAJM masih bergantung pada ISP dimana akses terhadap website yang dilakukan oleh pengguna jaringan akan selalu diteruskan ke ISP untuk di filter, tetapi ketika pengguna mengakses situs atau mencari kata kunci negatif yang belum terupdate dari ISP maka

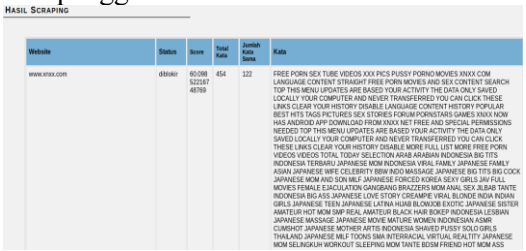
pengguna tetap bisa membuka situs tersebut. Sedangkan sistem keamanan yang dibuat oleh peneliti dengan metode *text filtering* mampu memfilter website apapun yang diakses oleh pengguna jaringan dan langsung mengklasifikasikan website tersebut layak atau tidaknya untuk diakses.

Proxy server yang terpasang pada topologi jaringan di UAJM akan menerima semua akses terhadap website yang diminta oleh pengguna jaringan kemudian memberikan url valid kepada sistem untuk diolah. Metode text filtering yang digunakan akan mengambil semua isi teks dari website, kemudian akan membandingkan isi teks tersebut dengan data training menggunakan build-in package dari bahasa program python yaitu regex yang menggunakan algoritma recursive backtracking. Dari hasil perbandingan menggunakan algoritma recursive backtraking sudah dapat ditentukan apakah website tersebut layak untuk diakses bagi pengguna jaringan di UAJM atau tidak.



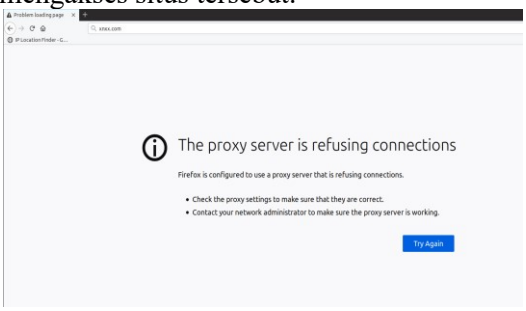
Gambar 16. Pengguna Mengakses Situs Negatif

Setelah proxy server menerima data bahwa ada permintaan atas sebuah situs, maka data tersebut akan diteruskan ke sistem dan diolah, kemudian sistem akan melakukan proses perhitungan terhadap tingkat kecocokan kata menggunakan algoritma *recursive backtracking* antara kata yang di *scrap* dengan data *training* yang terdapat di *database*. Jika situs tersebut memiliki tingkat kecocokan 10% untuk website radikalisme dan 30% utuk website pornografi atau lebih maka website tersebut akan dikategorikan sebagai website terblokir, jika sebaliknya maka website tersebut aman untuk diakses oleh pengguna.



Gambar 17. Website Terblokir

Dari gambar 17, dapat dilihat situs negatif yang diakses pengguna memiliki tingkat kecocokan 60% maka setelah itu situs tersebut akan dinyatakan terblokir, maka sistem akan mengupdate ACL dari proxy server dan menambahkan url dari situs yang sebelumnya diakses ke file ACL dan pengguna jaringan tidak dapat lagi mengakses situs tersebut.



Gambar 18. Situs Tidak Dapat Diakses

5. KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan dapat diambil kesimpulan sebagai berikut:

1. Sistem keamanan akses internet berbasis *text filtering* yang dirancang telah dapat menganalisis isi konten teks dari sebuah halaman website yang diakses pengguna dalam lingkungan UAJM.
2. Sistem yang dirancang telah dapat membatasi pengguna dalam mengakses website pornografi dan radikalisme dalam lingkungan UAJM, dengan mengklasifikasikan dan memblokir website – website tersebut.

6. DAFTAR PUSTAKA

[1] Arjuni.S, 2010. Perancangan dan Implementasi Proxy Server dan Manajemen Bandwith Menggunakan Linux Ubuntu Server Studi Kasus di Kantor Manajemen PT. Wisma Bumiputera Bandung. Tugas Akhir. Tidak diterbitkan. Institut Teknologi Bandung.

[2] Firmansyah & riadi. 2014. Analisis dan Perancangan Proxy Server Menggunakan Virtual Machine. e-ISSN: 2338-5197 Vol. 2 No. 3. Oktober 2014.

[3] Saurkar, Pathare, Gode. 2018. *An Overview On Web Scraping Techniques And Tools International Journal on*

Future Revolution in Computer Science & Communication Engineering Volume: 4 Issue: 4. ISSN: 2454-4248.

[4] Gosset, Sorter 2011. *Effectivenes of Internet Content Filtering*. Journal of Information Technologies Impact Vol. 11 No. 2. 2011.

[5] R. Rohimah. 2017. Kontribusi guru pendidikan agama Islam dalam menangkal potensi paham radikalisme (studi kasus di SMK Negeri 4 Semarang). UIN Walisongo.

[6] Muh.Subhan, A. Sudarsono, A. Barakbah, 2017. *Classification of Radical Web Content in Indonesia using Web Content Mining and k-Nearest Neighbor Algorithm*. EMITTER International Journal of Engineering Technology. Vol. 5, No. 2.

[7] Lubis, S.Husen. 2018. Membangun Korpus Pornografi Bahasa Indonesia. USU Sumatera.

[8] Ma'arif, Muhammad Rifqi, 2016. Integrasi Laman Web Tentang Pariwisata Daerah Istimewa Yogyakarta Memanfaatkan Teknologi *Web Scraping dan Text Mining*. TEKNOMATIKA 9.1 (2016): 71-80.

[9] Vargiu, 2013. *Exploiting web scraping in a collaborative filtering- based approach to web advertising*. Artificial Intelligence Research, 2013, Vol. 2, No. 1.