

Aplikasi Berbasis Web dengan Metode Crawling sebagai Cara Pengumpulan Data untuk Mengambil Keputusan

FA Suharno¹, L. Listiyoko²

¹Magister Teknik Informatika, STMIK Raharja, Tangerang
Jl. MH Thamrin, Kota Modernland, Tangerang

²Sistem Informasi, STMIK Muhammadiyah Banten, Tangerang
Jl. Syech Nawawi, Matagara, Tigaraksa, Tangerang

E-mail: frans.suharno@raharja.info 1), langgeng.listiyoko@stmikmbanten.ac.id 2)

ABSTRAKS

Di era modern, orang dapat saling bertukar informasi satu sama lain dengan mudah. Itu bisa berasal dari berbagai platform seperti media sosial, situs web, dan media online lainnya yang terhubung dengan orang, sehingga masalah keamanan akan menjadi sangat penting. Jika seseorang memposting satu informasi yang tidak dipercaya, maka kekacauan pun akan hadir. Ini sangat berbahaya untuk mengatur stabilitas, oleh karena itu aplikasi berbasis web untuk menganalisis dibuat. Aplikasi ini dibangun dengan metode Crawling, mengenai pengambilan keputusan apakah informasinya benar atau tidak. Aktifitas ini mendekati seperti pada Teks Mining.

Kata Kunci: *Crawling, Social Media, Aplikasi, Web Based Application, Text Mining.*

1. PENDAHULUAN

1.1. Latar Belakang

Pertukaran informasi yang sangat mudah dilakukan dewasa ini menimbulkan masalah baru dalam hal keamanan dan stabilitas nasional. Dengan kebebasan menyampaikan pesan melalui media sosial menyebabkan peluang penyebaran berita bohong dan berakibat pada timbulnya keresahan dalam masyarakat yang secara akumulatif bahkan berpotensi terjadi perpecahan. Dalam tahun-tahun politik menjelang pemilihan umum misalnya, sangat berbahaya jika tidak ada pengendalian tentang arus informasi yang memanfaatkan media sosial.

Institusi pertahanan sebagai pengemban tanggung jawab stabilitas nasional memerlukan sistem monitoring yang membantu kegiatan penyelidikan untuk dalam rangka mendapatkan limpahan informasi yang otentik dari media sosial dan mengubahnya menjadi bahan pertimbangan bersifat intelijen yang dapat ditindaklanjuti. Sistem akan bekerja berdasarkan informasi yang didapat dari sebuah modul yang dapat menghimpun dan melakukan verifikasi untuk memastikan kebenaran informasi. Pengelolaan informasi dimaksud memanfaatkan teknik text mining yang dibangun berbasis web.

Adapun modul data retrieval sebagai robot pekerjaannya dibangun sebuah aplikasi web based yang bekerja menggunakan teknik crawling. Dengan demikian sejumlah besar informasi yang dibutuhkan dapat dihimpun dengan efektif untuk selanjutnya dilakukan mining terhadapnya dan digunakan sebagai dasar pengambilan keputusan.

Business Intelligence (BI) memberikan kemampuan untuk menganalisis informasi dalam pengambilan keputusan manajemen pimpinan yang secara khusus dalam hal ini adalah kegiatan analisis intelijen. BI selanjutnya berperan juga dalam mengukur kemampuan aplikasi sejauh mana dapat diimplementasikan sehingga akurasi informasi semakin meningkat.

1.2. Referensi

1.2.1 Metode Crawling

Aplikasi data retrieval berbasis Web dengan metode crawler atau yang dikenal juga web spider atau web robot (Udapure, Kale, & Dharmik, 2014) adalah suatu program yang dibangun dan dirancang dengan metode tertentu yang secara otomatis mengumpulkan semua data informasi yang diinginkan yang ada dalam bermacam sumber website. Sebuah program yang melintasi struktur hypertext dari web, dimulai dari sebuah alamat awal (seed) dan

secara sekursif mengunjungi alamat web di dalam halaman web. Aplikasi Web crawler mengambil informasi pada website yang diberikan kepadanya, kemudian menyerap dan menyimpan semua data informasi yang terkandung didalam website tersebut. Setiap kali aplikasi mengunjungi sebuah website, maka secara otomatis akan merekam semua link yang ada di halaman yang dikunjunginya itu untuk kemudian dikunjungi lagi satu persatu.

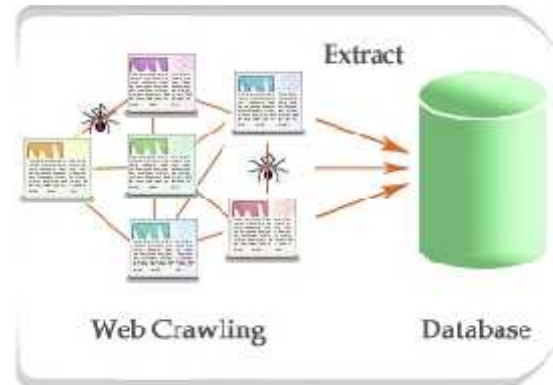
Selanjutnya aplikasi akan melakukan data retrieval dan menyimpannya ke dalam suatu media penyimpanan (*harddisk*) dengan kapasitas yang cukup besar. Data-data yang disimpan dalam hardisk ini kemudian, nantinya akan di ambil atau diakses pada saat dilakukan query (mining data). Dalam prosesnya Crawling data yang berhasil dihimpun dapat mencapai milyaran sementara penyajiannya dapat dilakukan secara real time.

Aplikasi crawling menyajikan informasi dan memberikan tampilan antarmuka dengan berbagai bentuk seperti hubungan, mode pencarian berdasarkan sosial media, tampilan statistik, statistik berbasis Tag Cloud, yang dipadukan dengan informasi yang dinamis dan relevan melalui media sosial.

1.2.2 Konsep Aplikasi

Proses mengunjungi website yang dilakukan crawler dimulai dari mengambil data dan informasi di semua seluruh URL dari website, menelusurinya satu-persatu, kemudian memasukkannya dalam daftar halaman pada indeks search engine, sehingga setiap kali ada perubahan pada website tersebut, akan terupdate secara otomatis. Aplikasi Web crawling juga melakukan proses mengambil dan mengumpulkan hyperlink yang ada dalam sebuah web yang dikunjunginya kemudian simpan dalam media berkapasitas besar. Indexing data yang berupa list hyperlink dilakukan demi memperlancar kegiatan penelusuran oleh search engine.

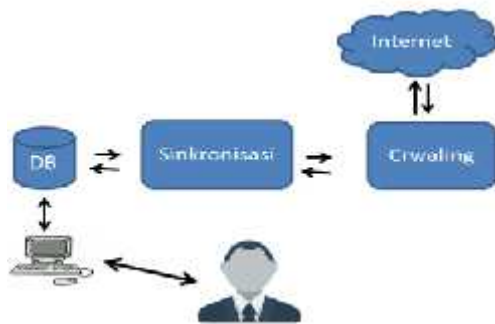
Aplikasi Web crawler termasuk kedalam bagian software agent atau yang lebih dikenal dengan istilah program bot. Secara umum sistem crawler ini memulai prosesnya dengan memberikan daftar sejumlah alamat website untuk dikunjungi, disebut sebagai seeds. Setiap kali sebuah halaman web dikunjungi, proses crawler ini akan mencari alamat yang lain yang terdapat didalamnya dan menambahkan kedalam daftar seeds sebelumnya.



Gambar 1. Pencarian Konsep Crawling

1.2.3 Manfaat Aplikasi

Perusahaan-perusahaan besar seperti Google dan Facebook telah berhasil memberdayakan Big Data untuk mengidentifikasi peluang-peluang bisnis, meningkatkan kualitas dan kuantitas layanan terhadap konsumennya, dan meningkatkan keuntungan. Google telah mengumpulkan data dalam volume sangat besar dengan cara crawling / mengorek / menyerap isi web page dari seluruh dunia. Dengan menggunakan data hasil crawling tersebut, yang dikombinasikan dengan data-data lain yang didapatkan dari kata-kata penelidikan (queries) yang diajukan via mesin pencariannya, Google telah berhasil meningkatkan efektifitas dari Google Adwords, dan dalam waktu yang sama juga meningkatkan kualitas pengalaman penggunaannya saat melakukan pencarian. Sebagai hasilnya, Google sukses meningkatkan keuntungan untuk dirinya sendiri dan juga untuk pelanggan pengguna layanan Adwords, serta mampu meningkatkan kepuasan para penggunaannya. Untuk dapat membangun bisnis berbasis data dalam skala sekelas Google, tentu membutuhkan waktu, tenaga, dan biaya yang tidak sedikit. Jangankan dalam skala besar, bahkan untuk memulai membuat suatu produk ataupun layanan berbasis data pada level pemula dan dalam skala kecil saja juga bukan hal yang mudah. Salah satu alasan utama kenapa begitu sulit untuk menghasilkan suatu produk berbasis data adalah karena terlebih dahulu kita mesti mampu mengumpulkan data dalam volume raksasa sebelum kita bisa membuat suatu produk yang berbasiskan data tersebut.



Gambar 2. Alur Crawling

Data atau informasi yang diinginkan dapat diperoleh dengan beberapa cara, diantaranya:

1. Input langsung "Kalimat" pada searching aplikasi, sehingga mesin pencari akan memproses dan disimpan dalam Data Base.
2. Dengan melakukan pencarian melalui Sumber informasi bisa di dapat melalui pihak ketiga seperti Facebook, Twitter, Instagram dan sebagainya.
3. Dengan melihat statistik kita dapat mengetahui tren yang berkembang saat ini.

Dari cara-cara pengumpulan data tersebut, sebagian besar perusahaan berbasis data pada kenyataannya menggunakan web crawler karena kebanyakan data yang dibutuhkan oleh perusahaan-perusahaan ini adalah dalam bentuk web page yang tidak bisa diakses menggunakan API tertentu. Namun demikian, meng-crawl web page bukanlah perkara mudah. Web adalah lautan informasi dengan miliaran web pages (laman web) dibuat setiap hari. Sebagian besar dari laman web tersebut memuat data-data yang tak berstruktur dan berantakan. Mengumpulkan dan mengorganisir data-data yang berantakan ini tentu bukan hal yang mudah dilakukan.

Di lingkungan institusi pemerintah khususnya TNI dan Polri sistem crawling ini mulai dikenalkan dan dimanfaatkan guna untuk mendapatkan informasi atau berita yang kebenarannya perlu di uji dan dianalisis apakah sesuai atau tidak dengan realita sebenarnya. Aplikasi yang digunakan salah satunya osint yaitu open source intelligence, aplikasi ini digunakan untuk tujuan penyelidikan kasus berita hoax, tetapi dalam penggunaannya diperlukan proses memperbaiki dan mempertajam teknik analisis karena bertujuan oknum di web sulit mencuri, menggunakan nama dan referensi palsu untuk akun mereka. Operator harus menggunakan beberapa teknik dengan terus-menerus memantau akun yang mencurigakan, yang secara khusus dikunci untuk mengumpulkan yang terkait yang cenderung

hampir miring, atau bereaksi dengan cara yang tidak proporsional sampai penutupan akun.

Secara statistik, analisis lalu lintas akun, dalam pengamatan, mengarah pada informasi investigasi berharga berikut ini:

1. Bukti kegiatan kriminal;
2. Geolokasi;
3. Map akun;
4. Keseluruhan jaringan kriminal

Aplikasi ini dapat mengungkap Profil sosial seseorang sehingga dapat diungkapkan oleh penyidik ketika dibutuhkan. Kemudian dilakukan pengidentifikasi seperti lingkaran keluarga, persahabatan, kota tempat dia tinggal, sekolah tempat dia tinggal, dimana dia berlibur pada musim tertentu yang sudah terlalui, apakah dia mempunyai anak, apa rencananya untuk akhir tahun, dan seterusnya. Kegiatan analisis yang ada pada aplikasi perlu dilakukan secara paralel dengan mengumpulkan informasi dari blog dan forum online, aplikasi komunikasi, media dan web.

2. PEMBAHASAN

Dalam proses ini, pengumpulan data dan informasi melalui Crawling yang telah disimpan dalam database (server) kemudian dilakukan query. Data diambil melalui media sosial seperti Facebook, Twiter dan Instagram sebagai penerapan metode yang dilakukan serupa untuk mengidentifikasi sasaran, tang Cloud, relationship folower maupun following dan menampilkan informasi lokasi tweets.

Kegiatan analisis tidak hanya digunakan oleh penyidik yang berhubungan ke cyber, tetapi juga oleh satuan-satuan yang berfungsi sebagai counter berita jika terjadi penyebaran berita HOAX yang menyudutkan institusi. Keamanan cyber pada masa sekarang banyak menggunakan teknik yang ada pada aplikasi untuk tujuan defensif dan pencegahan serangan berbahaya ke struktur strategis nasional khususnya di dunia militer, perbankan, perusahaan dan agen pemerintah.

a. Form Login

Halaman login diperlukan untuk mengatur hak akses user atau operator sehingga integritas dan keamanan sistem dapat tetap terjaga, seperti ditunjukkan dalam Gambar 3.



Gambar 3. Tampilan Login

b. Form Pencarian

Pemilihan kata kunci yang baik sangat disarankan demi menjaga privasi. Dengan pengaturan privilege sebagai common user akan didapat fitur informasi gambar, nama, Description, Start Time, End Time, Guest dan lokasi. Gambar 4 adalah tampilan halaman yang dapat difungsikan sebagai pencari informasi.



Gambar 4. Pencarian

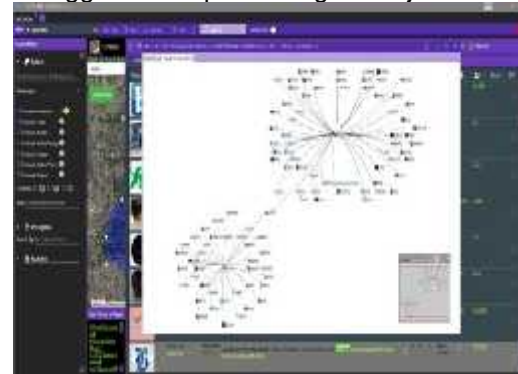
c. Form Pencarian Target

Salah satu fitur yang efektif adalah pencarian target berdasarkan lokasi seperti ilustrasi dalam Gambar 5. Dengan mode pencarian berdasarkan lokasi ini akan didapatkan informasi real yang diambil melalui post berdasarkan lokasi terakhir dilakukan.



Gambar 5. Pencarian Target

- c. Form Pencarian Mutual Teman Pada Twitter
Mode ini dengan memasukan username twitter pada kotak sehingga di temukan informasi Follower atau following pada twitter dalam bentuk Link Analisis. Gambar 6 menjelaskan informasi teman bersama (mutual friend) pengguna twitter dalam bentuk diagram. Analisis sederhana sudah dapat dilakukan secara manual berbekal informasi ini, namun jauh akan lebih efektif ketika dilanjutkan ke proses text mining menggunakan aplikasi tingkat lanjut.



Gambar 6. Pencarian Teman pada Twitter

3. KESIMPULAN

Dengan berkembangnya informasi teknologi dan kurangnya sarana pendukung keamanan data di mana penggunaan media sosial sangat diperlukan, setiap orang harus memiliki budaya dan oleh karena itu merupakan konsep aplikasi OSINT, baik untuk melakukan survei online untuk melindungi kepentingan mereka dan hanya untuk mengetahui bagaimana melindungi diri terhadap bahaya web. Pelanggaran dan pengetahuan terbaik dari musuh, pada tahun-tahun mendatang, kita akan mengembangkan Big Data sehingga perusahaan yang lebih memahami Big Data akan mampu menyuguhkan pengalaman yang lebih baik kepada para pelanggannya, mampu untuk menemukan pelanggan yang potensial, dan tentu mampu menghasilkan keuntungan yang lebih besar. Web crawling adalah bagian yang sangat penting dalam mengembangkan hampir semua produk ataupun layanan yang berbasis data.

PUSTAKA

- Anderson, D.W., Vault, V.D. & Dickson, C.E. 1999. *Problems and Prospects for the Decades Ahead: Competency Based Teacher Education*. Berkeley: McCutchan Publishing Co.
- Ary, D., Jacobs, L.C. & Razavieh, A. 1976. *Pengantar Penelitian Pendidikan*. Terjemahan oleh Arief Furchan. 1982. Surabaya: Usaha Nasional.
- Hitchcock, S., Carr, L. & Hall, W. 1996. *A Survey of STM Online Journals, 1990-1995: The Calm before the Storm*, (Online), (<http://journal.ecs.soton.ac.uk/survey/survey.html>, diakses 12 juni 1996).
- Jawa Pos. 22 April, 1995. *Wanita Kelas Bawah Lebih Mandiri*, hlm. 3.
- Kansil, C.L. 2002. Orientasi Baru Penyelenggara Pendidikan Program Profesional dalam Memenuhi Kebutuhan Dunia Industri. *Transpor*, XX (4): 57-61.
- Kumaidi. 1998. Pengukuran Bekal Awal belajar dan Pengembangan Tesnya. *Jurnal Ilmu Pendidikan*, (Online), Jilid 5, No. 4, (<http://www.malang.ac.id>, diakses 20 Januari 2000).
- Kuncoro, T. 1996. *Pengembangan Kurikulum Pelatihan Magang di STM Nasional Malang Jurusan Bangunan, Program Studi Bangunan Gedung: Suatu Studi Berdasarkan Kebutuhan Dunia Usaha Jasa Konstruksi*. Thesis tidak diterbitkan. Malang: PPS IKIP MALANG.
- Pitunov, B. 13 Desember, 2002. Sekolah Unggulan atukah Sekolah Pengunggulan? *Majalah Pos*, Hlm. 4 & 11.
- Pusat Pembinaan dan Pengembangan Bahasa. 1978. *Pedoman Penulisan Laporan Penelitian*. Jakarta: Depdikbud. *Undang-undang Republik Indonesia Nomor 2 tentang Sistem Pendidikan Nasional*. 1990. Jakarta: PT Armas Duta Jaya.
- Russel, T. 1999. An Alternative Conception: Representing Representation. Dalam P.J. Black & A. Lukas (Eds.), *Children's Informal Ideas in Science* (hlm. 62-84). London: Routledge.
- Saukah, A. & Waseso, M.G. (Eds.). 2002. *Menulis Artikel untuk Jurnal Ilmiah* (Edisi ke-4, cetakan ke-1). Malang: UM Press.
- Waseso, M.G. 2001. *Isi dan Format Jurnal Ilmiah*. Makalah disajikan dalam Seminar Lokakarya Penulisan Artikel dan Pengelolaan Jurnal Ilmiah, Universitas Lambungmangkurat, Banjarmasin, 9-11 Agustus.
- Wilson, D. 20 November 1995. Summary of Citing Internet Sites. *NETTRAIN Discussion List*, (Online), (NETTRAIN@ubvm.cc.buffalo.edu, diakses 22 November 1995).