

# STUDI DAN PENERAPAN ALGORITMA PENCARIAN MELEBAR (*BREADTH FIRST SEARCH*) PADA WEBSPIDERS DENGAN MENGUNAKAN APLIKASI *TELEPORT PRO*

Rhesa Adyithia – NIM : 135 05 081

Program Studi Teknik Informatika  
Sekolah Teknik Elektro dan Informatika, Institut Teknologi Bandung  
Jl. Ganesha 10 , Bandung  
e-mail: [if15081@students.if.itb.ac.id](mailto:if15081@students.if.itb.ac.id)

## ABSTRAK

Makalah ini membahas tentang studi dan penerapan algoritma Pencarian Melebar (*Breadth First Search*) pada aplikasi Teleport Pro untuk melakukan *Webspiders*. Algoritma Pencarian Melebar (*Breadth First Search*) adalah suatu algoritma pencarian graf yang dimulai dari suatu simpul akar dan mengunjungi semua simpul yang bertetangga dengan simpul akar. Kemudian, untuk setiap simpul terdekat mengunjungi simpul-simpul yang belum pernah dikunjungi. Sampai semua simpul telah dikunjungi. Dalam penerapannya di mesin *WebSpiders*, pencarian ini dilakukan sampai suatu kedalaman yang ditentukan. *WebSpiders* adalah suatu program atau script yang bekerja secara otomatis untuk menjelajahi *World Wide Web* dengan cara yang metodologis dan otomatis. Secara umum, program atau script dari *Webspiders* memulai aksinya dengan suatu daftar URLs yang akan dikunjungi, daftar ini disebut *seed*, kemudian mengidentifikasi setiap *hyperlinks* pada halaman tersebut. Setiap *hyperlink* yang teridentifikasi dimasukkan dalam suatu daftar yang disebut *crawl frontier*. Selanjutnya secara rekursif, URLs dari *crawl frontier* tersebut dikunjungi. Proses pencarian oleh *Webspiders* ini dapat dilakukan dengan dua algoritma, salah satunya Algoritma pencarian melebar. Penggunaan metode ini menyebabkan *webspiders* menyelesaikan seluruh proses identifikasi *hyperlink* dari satu URL sebelum menuju URL selanjutnya yang terdapat dalam satu antrian. Teleport Pro adalah salah satu aplikasi *webspiders* yang menggunakan algoritma Pencarian Melebar (*Breadth First Search*) dalam melakukan tugasnya. Teleport pro memberikan hasil yang cukup baik dalam mengunduh keseluruhan dari suatu website sehingga penggunaanya dapat mengakses keseluruhan website secara *offline*. Aplikasi Teleport Pro mulai dikembangkan sebelum tahun 1997 dan masih terus dikembangkan sampai sekarang.

**Kata kunci:** Algoritma, *Algorithm*, *Graph Search Algorithm*, Algoritma Pencarian Graf, Pencarian Melebar, *Breadth First Search*, BFS, *Webspiders*, Teleport Pro.

## 1. PENDAHULUAN

Perkembangan teknologi informasi yang begitu cepat memacu kebutuhan akan Informasi. Dalam dunia digital, Informasi yang dibutuhkan berupa data. Data ini dapat berupa gambar, musik, video, dokumen, dan sebagainya. Salah satu sumber informasi terbesar adalah internet. Dalam dunia maya yang disebut internet ini, seseorang dapat mengakses informasi apa saja yang dibutuhkan jika tersedia di suatu tempat dalam internet. Seseorang yang membutuhkan suatu informasi harus mengetahui alamat dari informasi tersebut. Pencarian alamat dari informasi yang dibutuhkan tersebut biasanya dilakukan oleh suatu mesin pencari (*search engine*). Mesin pencari sering disebut sebagai kunci dari informasi spesifik agar dapat ditemukan di internet. Tanpa mesin pencari, hampir tidak mungkin suatu informasi yang sesuai dengan keinginan pengguna dapat ditemukan.

Sebuah mesin pencari menggunakan metode atau algoritma dalam pencarian informasi. Algoritma yang umum dipakai dalam mesin pencari tersebut adalah Algoritma pencarian mendalam (*Depth First Search*) dan Algoritma pencarian melebar (*Breadth First Search*). Secara umum suatu mesin pencari memulai aksinya dari suatu titik yang diketahui, mencari semua *hyperlink* yang ada, dan meneruskan ke titik selanjutnya dalam antrian. Program atau *Automated script* dari mesin pencari ini sering disebut dengan *WebSpiders* atau *WebCrawlers*.

Adanya kebutuhan akan informasi yang dapat diakses kapan saja melalui komputer yang tidak terhubung dengan internet menjadi salah satu penyebab utama dari *data mining*. *Data mining* adalah suatu kegiatan pengumpulan data dari internet sehingga seseorang dapat mengakses data tersebut tanpa terhubung dengan sumber aslinya di

internet, tetapi hanya mengakses data yang sudah diunduh pada komputernya.

Ada banyak aplikasi untuk melakukan *data mining*. Salah satu aplikasi yang cukup handal dalam melakukan data mining adalah Teleport Pro. Aplikasi ini dapat melakukan *data mining* terhadap suatu website secara menyeluruh sehingga pengguna dapat seakan-akan mengakses website tersebut di komputernya sendiri. Aplikasi ini mengubah seluruh *hyperlink* yang terdapat dalam suatu halaman web ke halaman web yang terdapat di komputer yang bersesuaian dengan halaman asli di website aslinya.

## 2. METODE

Metode yang dibahas dalam bagian ini adalah metode dari Webspiders yang menggunakan Algoritma pencarian melebar (*Breadth First Search*) dan metode tersendiri dari suatu *data mining*.

### 2.1 Webspiders dengan *Breadth First Search*

Webspiders dengan algoritma pencarian melebar (*Breadth First Search*) mengidentifikasi seluruh hyperlink dalam satu halaman web dahulu sebelum melanjutkan ke halaman web selanjutnya.

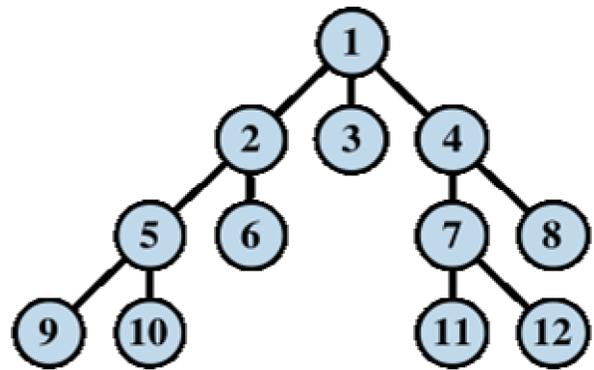
Mekanisme kerja dari Webspiders dengan algoritma pencarian melebar (*Breadth First Search*) adalah sebagai berikut :

1. Mengunjungi suatu halaman web yang ada dalam antrian
2. Membaca dan menyimpan informasi yang ada dalam web tersebut
3. Membaca *Meta Tag* dari website
4. Mencari seluruh hyperlink yang ada di website tersebut dan memasukkannya dalam antrian
5. Melakukan langkah 1.

Dalam melakukan pencariannya, suatu Webspiders terus menerus melakukan langkah-langkah diatas tanpa batasan kedalaman.

*Webspiders* memulai aksinya dengan suatu daftar URLs yang akan dikunjungi, daftar ini disebut *seed*. Dari setiap halaman website yang telah dibuka, *Webspiders* mencari setiap *link* ke halaman lainnya. Daftar yang selanjutnya diekspansi disebut sebagai *Crawl Frotier*. *Crawl Frontier* inilah yang disebut sebagai antrian.

Dengan menganalisis hubungan keterkaitan antar halaman web, suatu *Webspiders* dapat melihat abstraksi dari isi halaman web tersebut. Selain itu *Webspiders* juga dapat menentukan "*Importance level*" dari suatu halaman web dan urutannya.



Gambar 1. Urutan dari simpul (halaman web) yang di ekspansi oleh Webspiders

*Webspiders* dengan algoritma pencarian melebar (*Breadth First Search*) melakukan tugasnya dengan baik pada *data mining*, khususnya pada penggunaannya dalam aplikasi Teleport Pro. *Webspiders* ini mengidentifikasi seluruh *link* pada satu halaman web dahulu (yang terdapat pada antrian) sebelum melanjutkan pencariannya pada halaman web selanjutnya.

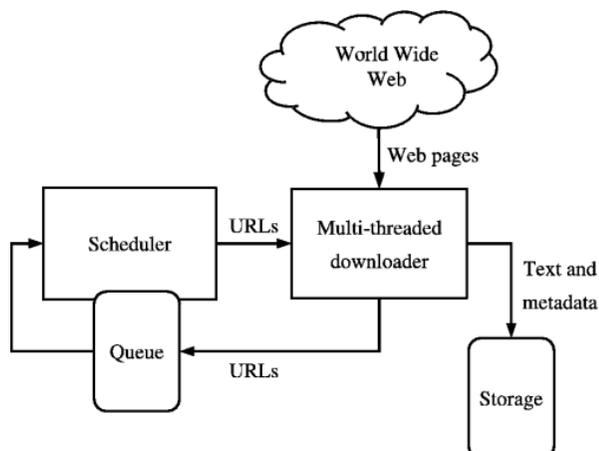
Metode ini memberi keunggulan dalam proses *data mining* karena satu halaman web sudah dipastikan disimpan dan seluruh *link* yang terdapat di dalamnya sudah diketahui dan masuk dalam suatu antrian sebelum *Webspiders* menelusuri halaman web selanjutnya sehingga tidak diperlukan akses berulang kali pada satu halaman web.

Penggunaan metode ini juga memastikan tidak diperlukannya penyimpanan posisi pembacaan terakhir dalam satu halaman web ketika membaca halaman web selanjutnya.

Selain memiliki antrian, *Webspiders* harus memiliki suatu daftar yang berisi indeks dari halaman web yang sudah dikunjungi. Daftar ini memiliki fungsi untuk mencegah pengaksesan berulang ketika halaman web tersebut memiliki *hyperlink* kembali ke halaman web sebelumnya atau yang sudah pernah diakses. Atau yang sering disebut dengan *Spider Trap*. Secara rinci *Spider Trap* dijelaskan sebagai suatu kondisi dinamis dari suatu website sehingga *Webspiders* terjebak dalam pencarian berulang kode yang tidak pernah berakhir.

Kinerja dari suatu *Webspiders* tidak hanya ditentukan oleh strategi pencarian dan pengambilan data saja. Desain arsitektural yang dioptimasi dengan baik menjadi faktor lain yang sangat berpengaruh dengan kinerja dari suatu *Webspiders*.

Gambar 2 menunjukkan salah satu desain arsitektural dari *Webspiders* yang biasa digunakan.



**Gambar 2. Desain arsitektural dari Web spiders**

Secara arsitektural, suatu *web spiders* terdiri dari bagian seperti dalam gambar 2. *Multi-threaded downloader* bekerja sebagai pengunduh data dan juga pencari *hyperlink* yang terdapat dalam satu halaman web. Data yang telah diunduh disimpan pada *storage*, sedangkan alamat *hyperlink* diberikan pada suatu antrian, yang kemudian dikendalikan oleh *Scheduler*, kapan alamat tersebut diberikan pada *Multi-threaded downloader* untuk mengunduh halaman web selanjutnya.

Dalam *Web spiders* dengan arsitektural umum, suatu Algoritma yang dipakai menyebabkan perubahan pada sistem pencarian *hyperlink* dari *Multi-threaded downloader* yang menentukan masukan alamat halaman web pada antrian juga pada bagian *Scheduler* yang mengatur pengunduhan halaman web dalam antrian oleh *Multi-threaded downloader*.

Pada umumnya, *Web spiders* yang dipakai dalam suatu aplikasi ataupun mesin pencari di internet sudah dimodifikasi sedemikian rupa oleh masing-masing pembuatnya. Sehingga desain arsitektural dan penggunaan algoritma yang rinci pada setiap aplikasi atau mesin pencari di internet tidak dipublikasikan dengan bebas.

## 2.2 Metode Data Mining

*Data mining* adalah suatu proses dari penemuan suatu data atau informasi yang diperlukan dari sejumlah besar data atau informasi yang lain dengan cara yang mangkus.

Proses kerja dari *Data mining* memberikan perhatian pada deteksi secara otomatis dari fakta-fakta dan hubungan keterkaitan dari data-data tersebut. Proses *Data mining* terdiri dari 6 kegiatan yaitu klasifikasi, pendekatan, perkiraan, penggabungan, pengklusteran, dan deskripsi dan visualisasi.

Secara garis besar, ada 3 proses utama yaitu Persiapan data yang akan digunakan atau diperlukan untuk mendapatkan hasil, Penganganan data yang lebih lanjut untuk memudahkan pencarian informasi, dan yang

terakhir Analisis data sehingga dihasilkan data yang sesuai dengan keinginan.

Proses analisis dari suatu *data mining* menggunakan banyak metode. Metode tersebut mulai dari pendekatan berbasis kasus, pemikiran manusia, sampai ke penggunaan *Intelligent Agents*.

Pada *Web spiders*, Proses *data mining* dengan basis *Intelligent Agents* sangat diperlukan. Data yang harus diolah adalah suatu halaman web. Dari sekian banyak teks yang terdapat dalam halaman web, harus diidentifikasi alamat mana saja yang mengacu pada halaman web yang lain, ataupun sumber data yang lain diluar teks yang terdapat dari halaman tersebut seperti data yang berupa gambar, musik, film dan sebagainya.

Hasil identifikasi tersebut masuk ke dalam proses selanjutnya yaitu pengelompokan informasi (data yang sudah diproses) berdasarkan jenisnya, seperti alamat dari suatu halaman web lain, maupun alamat dari suatu data yang berupa gambar pada *file* terpisah dari halaman web yang menjadi sumber masukan data.

Setelah pengelompokan setiap data yang sudah dikelompokkan masuk ke tahap analisis. Analisis ini mencari alamat yang relevan dengan apa yang dicari. Baik melalui pembacaan meta (untuk suatu halaman web), maupun nama dari alamat yang berasal dari proses sebelumnya.

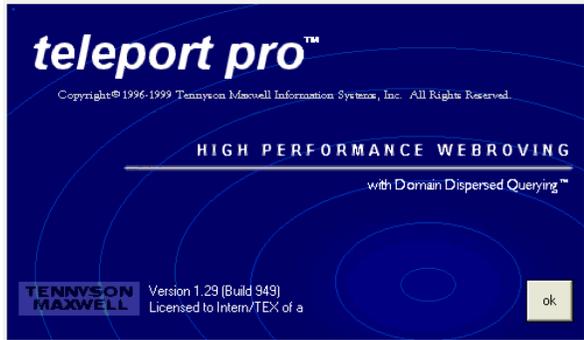
Hasil analisis dari proses *data mining* ini akan dilemparkan kembali pada suatu antrian, atau suatu bagian lain dari *web spiders* yang mengelola urutan dari data yang masih harus diambil, maupun lokasi dari data tersebut.

## 3. TELEPORT PRO

Teleport Pro adalah salah satu aplikasi *web spiders* yang menggunakan algoritma Pencarian Melebar (*Breadth First Search*) dalam melakukan tugasnya. Teleport Pro sudah dikembangkan sebelum tahun 1997 dan masih terus dikembangkan sampai sekarang. Teleport Pro adalah aplikasi yang dibuat oleh *Tennyson Maxwell Information System, Inc.*

Teleport Pro bekerja seperti layaknya sebuah *browser* yang menerima *file html*. Setiap *hyperlink* yang terdeteksi oleh *Web spiders* dalam suatu halaman web, diklasifikasikan menjadi *file* yang dapat memiliki *hyperlink* didalamnya dan *file* yang tidak mungkin memiliki *hyperlink* dengan alamat lainnya. Setiap hubungan dengan file diluar file tersebut masuk dalam sat daftar antrian. Setiap file yang masuk dalam katagori *file* yang tidak mungkin memiliki *hyperlink* seperti *file* gambar, musik, flash (dianggap tidak memiliki *hyperlink* ke tempat lain) diunduh terlebih dahulu. Setelah semuanya diunduh, file yang dapat memiliki *hyperlink* mulai diunduh. Ketika proses pengunduhan dilakukan, satu per satu file yang sedang diunduh diperiksa *Web spiders*, apakah ada file lainnya, maupun *hyperlink* ke alamat lain

yang belum dikunjungi atau terdapat dalam antrian. Proses dilakukan terus menerus sampai suatu kedalaman yang telah ditentukan tercapai, atau semua alamat dari halaman web yang ada *hyperlink*nya sudah diunduh.



Gambar 3. Aplikasi Teleport Pro yang digunakan

Teleport Pro menggunakan *Webspiders* yang memiliki kapasitas sampai 10 buah *Downloader* yang dapat bekerja dalam waktu yang bersamaan. Basis data yang dapat dikelola oleh Teleport Pro hanya sampai 65.500 alamat halaman Web (alamat *file* yang lain tidak termasuk).

### 3.1 Pemakaian Teleport Pro

Pada Uji coba dari pemakaian Aplikasi Teleport Pro ini, digunakan Aplikasi Teleport Pro versi 1.29.949 yang dirilis pada bulan Februari tahun 1999. Aplikasi Telpert Pro yang digunakan versi Trial yang dapat dipakai sampai 40 kali pengunduhan secara gratis. Uji coba pengunduhan website dilakukan pada website dengan alamat : <http://kur2003.if.itb.ac.id> yang berada dalam lingkungan intranet ITB.

Pengujian dilakukan pertama-tama dengan memilih fasilitas menu *New Project Wizard* yang mengkonfigurasi otomatis suatu proyek pengunduhan website pada Aplikasi Teleport Pro.



Gambar 4. Menu Langkah 1, memilih pengaturan

Pada Gambar 4 tampak bahwa Aplikasi Teleport Pro memiliki beberapa fitur, dari mencari website dengan suatu kata kunci, sampai membuat duplikat dari suatu website secara utuh termasuk struktur direktorinya. Dalam pengujian dipilih menu duplikat website dimana *Webspiders* yang baik harus digunakan dengan tepat.



Gambar 5. Menu Langkah 2, memasukkan alamat mulai

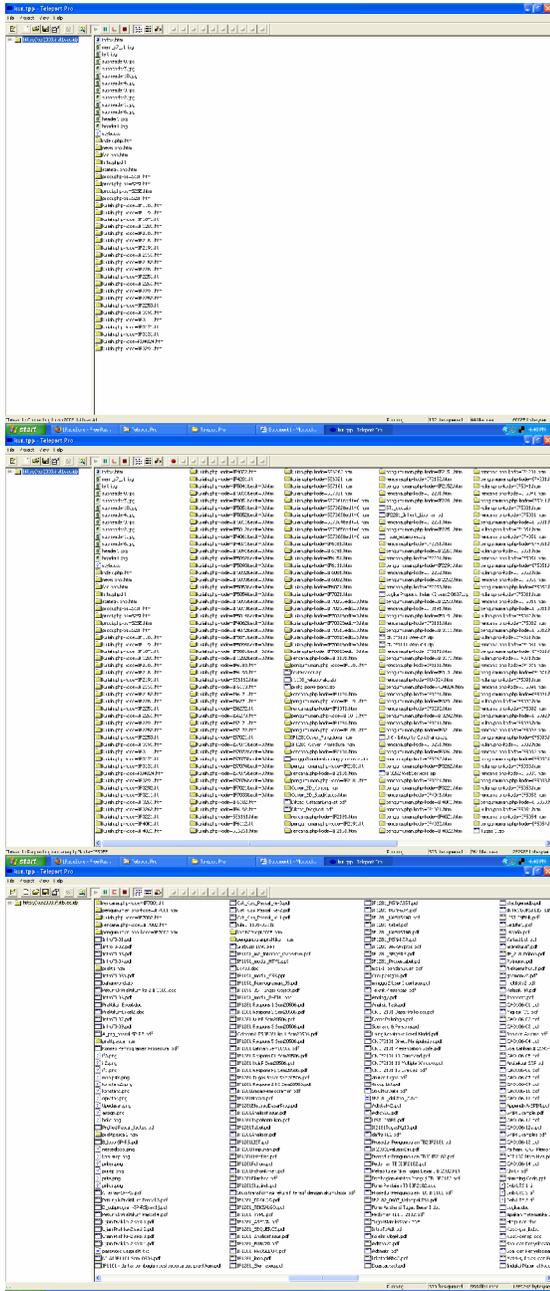
Alamat Mulai dari pengunduhan website harus dimasukkan setelah langkah 1. Dalam pengujian dipilih website dengan alamat <http://kur2003.if.itb.ac.id> (Lihat gambar 5)



Gambar 6. Menu Langkah 3, memilih pengaturan tipe *file* yang akan diunduh

Langkah selanjutnya adalah Pemilihan konfigurasi dari proyek (Gamar 6). Teleport Pro dapat di atur untuk mengunduh hanya teks, teks dan gambar, sampai seluruh isi dari website tersebut. Untuk menguji kinerja dari Algoritma Pencarian melebar, dipilih pengaturan pengunduhan semuanya.

Segera setelah semua pengaturan sudah dipilih, user diminta memasukkan nama proyek dan lokasi penyimpanannya file yang berisi keterangan mengenai proyek tersebut. Setelah semua pengaturan selesai, user dapat memulai pekerjaan untuk mengunduh suatu website sesuai proyek dengan menekan tombol start pada menu utama.



**Gambar 7. Teleport pro bekerja mengunduh keseluruhan isi website, Atas : Awal mulai, Tengah : Pertengahan, Bawah : Pada waktu akhir mendekati selesainya Pengunduhan suatu website**

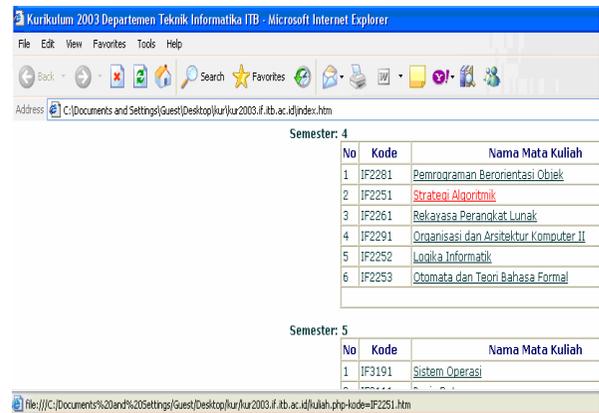
Aplikasi Teleport Pro akan memulai aksi pengunduhan websitenya. Proses dari berjalannya aksi ini dapat dilihat di gambar 7. Dari Proses pengunduhan ini dapat kita lihat bahwa Algoritma penelusuran (pengunduhan) website yang digunakan adalah Algoritma Pencarian Melebar (*Breadth First Search*).



**Gambar 8. Konfirmasi Proyek yang dijalankan telah selesai dan status akhir dari proyek (jumlah file yang dibaca dan disimpan dalam storage)**

Setelah Aplikasi selesai menjalankan sebuah proyek, Aplikasi Teleport Pro akan memunculkan sebuah kotak pemberitahuan yang mengidentifikasi bahwa proyek telah berhasil dan jumlah file yang dibaca (seluruh hubungan dengan file diluar file utama) dan jumlah file yang berhasil disimpan di storage.

Teleport Pro secara otomatis akan merubah seluruh link dengan file lain dengan alamat relatif dari file tersebut berada, sehingga dapat memudahkan siapapun menerima kumpulan file hasil unduh dari suatu website hanya dengan menyalin data tersebut ke tempat lain.



**Gambar 9. Contoh Website hasil pengunduhan oleh Teleport Pro**

Gambar 9 menunjukkan hasil pengunduhan dari website dengan alamat <http://kur2003.if.itb.ac.id>. Hasil pengunduhan suatu website menyerupai website aslinya pada waktu website tersebut diunduh. Setiap link maupun hyperlink yang terdapat dalam halaman utama tersebut sudah dirubah secara otomatis oleh Aplikasi Teleport Pro sesuai dengan alamat relatif dari file yang diunduh Teleport Pro. Teleport Pro memungkinkan peremajaan website hasil unduhan dengan memeriksa apakah ada

perubahan antara website hasil unduhan yang ada di storage dengan website asli.

### 3.2 Analisis Teleport Pro

Berdasarkan hasil pengujian Aplikasi Teleport Pro seperti dikemukakan pada bagian 3.1, dapat dikaji bahwa Aplikasi ini berjalan cukup baik dengan kecepatan transfer mencapai 1,3 MegaByte per Second untuk pengunduhan terhadap suatu website di intranet ITB dari lingkungan ITB sendiri.

Proses keseluruhan dari pengunduhan website tersebut memakan waktu kurang lebih 5 menit. Website yang diunduh memiliki besar 353 MegaByte. Dalam Website tersebut Teleport Pro mendeteksi ada 1383 *file* keseluruhan dan Teleport Pro dapat mengunduh 1328 *file*. Teleport Pro mencatat angka sampai 96% *file* yang dapat diunduh dari suatu website.

## 4. KESIMPULAN

Teleport Pro adalah suatu Aplikasi *Webspiders* yang menggunakan Algoritma Pencarian Melebar (*Breadth First Search*) dalam melakukan *Data Mining*. Teleport Pro memberikan hasil yang memuaskan dengan mencatat 96% *file* yang dapat diunduh berdasarkan hasil pengujian.

Teleport Pro mengunduh *file* yang tidak dapat memiliki hubungan dengan *file* lain sebelum mengunduh *file* yang mungkin dapat memiliki hubungan dengan *file* lain. Setiap *file* yang diunduh dimasukkan dalam suatu basis data yang mencatat *file* yang telah diunduh, alamat sumber, dan alamat file di *storage*. Basis data ini digunakan untuk meremajakan *link* yang terdapat pada setiap halaman web setelah proses pengunduhan suatu website selesai.

Proses Data Extracting oleh *Webspiders* pada Aplikasi Teleport Pro menunjukkan hasil yang memuaskan. Hal ini merupakan salah satu bukti keunggulan Algoritma Breadth First Search dalam proses *Data mining* (Pengunduhan) suatu website secara menyeluruh.

## REFERENSI

- [1] Reitano Antonella, "Data Mining and Management Decisions", *International Conference on Computer System and Technologies. CompSysTech '2005*, 2005.
- [2] Munir Rinaldi, "Diktat Kuliah IF2251 Strategi Algoritmik", Program studi Teknik Informatika, 2007.
- [3] Tenmax, Teleport Pro, <http://www.tenmax.com/teleport/home.htm>, Tanggal akses: 20 April 2007 pukul 15:00.

- [4] Wikipedia, WebSpiders, [http://en.wikipedia.org/wiki/Web\\_crawler](http://en.wikipedia.org/wiki/Web_crawler), Tanggal akses: 23 April 2007 pukul 14:00
- [5] Wikipedia. Breadth First Search. [http://en.wikipedia.org/wiki/Breadth-first\\_search](http://en.wikipedia.org/wiki/Breadth-first_search) Tanggal akses: 23 April 2007 pukul 14:00