

# WWW Mirroring dengan w3mir

Anda mungkin menemukan *website* yang menarik dan tidak selalu bisa *online* untuk membaca isinya. Bagi Anda yang ingin membuat *mirror* atau *local copy* untuk dapat dibaca lokal, Anda bisa mempergunakan w3mir. W3mir adalah *mirroring tool* yang *powerful*, namun mudah digunakan.

Terkadang, kita menemukan informasi yang begitu menarik di *www*, namun tidak tersedia *link* untuk men-download kumpulan informasi tersebut. Lebih parah lagi apabila kumpulan informasi tersebut tersebar dalam puluhan file, dimana akan menyulitkan apabila harus disimpan satu per satu. Kalaupun ingin menyimpannya, kita mungkin kehilangan *link* yang menghubungkan file-file yang kita download tersebut. Dalam kondisi demikian, beberapa pihak akan memilih cara untuk men-download keseluruhan kumpulan informasi tersebut agar dapat dibaca lokal dengan bantuan *www mirroring tool*.

Di dunia *open source*, kita mengenal begitu banyak program seperti ini. Salah satu dari tool tersebut yang cukup menarik untuk digunakan adalah w3mir, sebuah *www mirroring tool* yang cukup kaya fitur. W3mir dapat men-download keseluruhan direktori ataupun sekadar file-file tertentu. Selain itu, ketika men-download, w3mir juga akan memperbaiki URL agar sesuai dengan kondisi *local copy* Anda sehingga *local browsing* tetap dapat dilakukan. W3mir mendukung HTML4, CSS (*partial support*), Java, ActiveX dan file-file PDF, di luar fitur-fitur lainnya. Di artikel kali ini, kita akan membahas cara-cara mirroring beserta tip menggunakan w3mir.

W3mir dibangun dengan Perl sehingga Anda akan membutuhkan Perl terinstal di sistem Anda (hampir semua distro umum sudah melakukannya). Selain itu, beberapa modul perl harus diinstal agar w3mir dapat bekerja. Anda bisa mendapatkan perl-URI, perl-libwww-perl, perl-HTML-parser di

CPAN atau di CDROM distro Anda. Pustaka-pustaka tersebut cukup umum. Installah dengan tool *packaging* distro Anda untuk lebih mudahnya. Setelah semua kebutuhan w3mir terpenuhi, Anda bisa men-download tool ini di website-nya, <http://langfeldt.net/w3mir/>.

Untuk menggunakan, Anda bisa mempergunakan file konfigurasi atau tidak sama sekali. Apabila tidak mempergunakan file konfigurasi, Anda akan selalu memberikan opsi tertentu ketika menjalankan program ini. Untuk mudahnya, kita tidak akan mempergunakan file konfigurasi dalam artikel ini. Apabila diinginkan, Anda bisa meng-copy-kan contoh file konfigurasi yang datang bersama paket ini ke home directory Anda dengan nama *.w3mir*. Contoh konfigurasi tersebut sudah *self-documented*.

Untuk melihat opsi yang tersedia, seperti biasa, berikan opsi *-h* pada saat menjalankan w3mir. Penerapan opsi-opsi tersebut akan kita bahas sama-sama dalam bagian contoh. Untuk mencoba w3mir, alangkah baiknya apabila Anda terhubung ke jaringan lokal yang menyediakan file sharing dari *www*, atau Anda dapat membangun web server sederhana untuk mencoba.

## Contoh-contoh penggunaan

### • Menampilkan daftar file yang akan di-download, tanpa men-download terlebih dahulu.

Kadang-kadang, kita ingin mengetahui file-file apa saja yang akan di-download dalam proses mirroring. Beberapa pengguna yang paranoia umumnya akan melakukan hal ini. Berikan opsi *-f* untuk menampilkan

daftar file seperti contoh berikut:

```
$ w3mir -f http://192.168.0.1/pub/
w3mir: http://192.168.0.1/robots.txt
w3mir: index.html, got and forgot it.
w3mir: That's all (2073+463 bytes of it).
```

Untuk lebih baik lagi dalam menampilkan URL, Anda dapat mengombinasikannya dengan opsi *-l*, seperti contoh berikut:

```
$ w3mir -l -f http://192.168.0.1/pub/
w3mir: http://192.168.0.1/robots.txt
w3mir: index.html, got and forgot it.
...
w3mir: That's all (2073+463 bytes of it).
```

Sebagai catatan, Anda harus menuliskan opsi *-l* dan *-f* secara terpisah. w3mir tidak mengenali *-l* dan *-f* sebagai *-lf* atau *-fl*. Catatan berikutnya adalah, walaupun tidak ada file yang akan di-download ketika Anda membuat list seperti ini, file untuk kebutuhan internal w3mir, yaitu *.redirs* dan *.referrers* akan dibuat. Anda bisa mengabaikan file ini.

Selain untuk kebutuhan audit keamanan, daftar file ini sebenarnya cukup berguna. Asumsikan Anda tidak memiliki koneksi Internet, namun rekan Anda selalu bersedia untuk men-download URL pesanan Anda. Hanya, beliau tidak mau re-

pot-repot mencari URL apa yang Anda inginkan di *search engine*. Beliau meminta *list* URL agar dapat dilewatkan ke script untuk men-download file-file yang bersangkutan. Dalam hal ini, Anda bisa mempergunakan w3mir untuk menghasilkan list (terutama kalau Anda ingin men-download informasi dari beberapa website sekaligus, dengan filter file tertentu).

#### ● Men-download file tunggal.

Untuk men-download file tunggal dengan w3mir, Anda tidak perlu memberikan opsi apapun. Cukup sebutkan saja URLnya, seperti contoh berikut:

```
$ w3mir http://192.168.0.1/pub/OS/SUSE91/INSTALL/INDEX.gz
```

```
w3mir: http://192.168.0.1/robots.txt
```

```
w3mir: INDEX.gz: 34162 bytes
```

```
w3mir: That's all (35205+549 bytes of it).
```

#### ● Menampilkan progress.

Untuk file besar, terkadang kita ingin me-

lihat progress download yang kita lakukan. Sekadar untuk memastikan semuanya bekerja dengan baik, misalnya. Anda bisa memberikan opsi `-c` kepada w3mir dalam menampilkan progress peng-copy-an file. Sayangnya, opsi ini tidak terlalu bisa diandalkan. Terkadang opsi ini bekerja, terkadang tidak bekerja. Jadi, hanya gunakan kalau benar-benar perlu.

#### ● Men-download direktori secara rekursif.

Ini lah yang umumnya kita inginkan. Untuk mendapatkan fasilitas ini, di mana w3mir akan mengambil semua file dalam direktori yang kita minta, berikan saja opsi `-r`, dan secara otomatis sebuah local copy akan tersedia di harddisk Anda. Contoh:

```
$ w3mir -r http://192.168.0.1/pub/Apps/Office/Kamus/KamusLain1/
```

```
w3mir: http://192.168.0.1/robots.txt
```

```
w3mir: index.html: 812 bytes
...
...
```

```
w3mir: ?C=N: 812 bytes
```

```
w3mir: That's all (2159676+2439 bytes of it).
```

Setelah selesai download, perhatikanlah direktori aktif Anda. File yang ada di dalam direktori di server secara otomatis akan tersimpan.

#### ● Mirror keseluruhan site.

Untuk melakukan mirror keseluruhan situs, cukup berikan saja opsi `-r` pada direktori tertinggi dari site yang ingin Anda mirror. Hanya, Anda perlu sedikit cermat dalam melakukan hal seperti ini. Pertama, Anda mungkin perlu membatasi kedalaman rekursi ketika w3mir men-download. Secara default, w3mir akan terus mengikuti link apabila Anda memberikan opsi `-r`. Hal ini bisa menjadi bermasalah ketika website bersangkutan memasang perangkap. Anda bisa mempergunakan file konfigurasi dengan direktif `Ignore`.

Sebagai contoh:

```
Ignore: /*/*/*/*/*/*/*/*
```

## Colocation dengan 3 Internet Backbone (USA, EROPA & SINGAPORE)

harga mulai dari Rp. 1.325.000,-



#### Anda mendapatkan :

- 128 Kbps Bandwidth Internasional
- Unlimited 100 Mbps IIX Bandwidth
- 8 Public IP Address
- Technical Support 24 x 7 x 365
- 99.9% Uptime Guarantee
- UPS Backup

Email : sales@saranadata.com

Telp : +62 (21) 526 9257

#### Business Services :

- Colocation Server
- Webhosting Colocation
- Dedicated Server
- Rack Solution
- Internet Service Solution
- Corp. Application Hosting
- System Mirroring & DRC
- Data Storage Services
- Custom Made Solution

#### Personal Services :

- Download Services
- E- File Storage
- E- Photo Album

Direktif ini akan membatasi sampai ke-dalam 4 level (3 garis miring pertama untuk internal w3mir apabila url yang diberikan telah memiliki 3 garis miring pertama seperti `http://192.168.0.1/`).

### ● Menampilkan isi file.

Bagaimana kalau Anda ingin sekadar menampilkan isi dari suatu file namun tidak ingin menyimpan filenya ke harddisk Anda? Ini mirip dengan perintah `cat`. Caranya, berikan saja opsi `-s` (diikuti oleh nama file lengkap) pada *command line*.

Isi dari file yang Anda tampilkan tersebut akan tampil pada `stdout`. Dengan pemahaman tersebut, Anda bisa menyimpan isi file yang ditampilkan tersebut ke file dengan nama lain. Cukup arahkan `stdout` Anda ke file lain seperti contoh berikut:

```
$ w3mir -s http://192.168.0.1/pub/ > a.html
w3mir: http://192.168.0.1/robots.txt
w3mir: index.htmlw3mir: That's all (2073+463 bytes of it).
```

Apabila Anda buka file `a.html` tersebut, maka isinya akan sama dengan apa yang seharusnya Anda lihat ketika mengarahkan *browser* ke alamat yang kita berikan kepada w3mir.

### ● Berpura-pura sebagai agent lain.

Setiap `http agent` yang kita miliki, entah *web browser*, *downloader*, dan lain sebagainya umumnya mengatur nama agent sesuai dengan nama programnya. Ketika kita menggunakan w3mir, nama agent yang dikirim ke web server adalah `w3mir/version-release-date`. Ketika kita ingin men-download website yang hanya menerima agent tertentu misalnya, kita perlu berpura-pura sebagai agent yang diizinkan. Caranya adalah dengan memberikan opsi `-agent` dan menyebutkan nama agent. Kita tentu bisa menuliskan apa yang kita inginkan. Sebagai contoh:

```
$ w3mir -agent 'Mozilla/8.1 (compatible; MSIE7.0; Windows 2005)' -f http://192.168.0.1/pub/
```

### ● Menampilkan URL yang tersimpan dalam dokumen.

W3mir dapat pula mengetahui URL apa

saja yang terdapat suatu dokumen. Anda mungkin ingin menampilkannya. Untuk itu, berikan opsi `-l` (L kecil) seperti contoh berikut ini:

```
$ w3mir -l http://192.168.0.1/pub/Intermezzo/Interview/Fosdem\2002\ -\ report.htm
w3mir: http://192.168.0.1/robots.txt
http://www.freehackers.org/
...
...
...
w3mir: That's all (24460+561 bytes of it).
```

### ● Memproses file secara batch.

Secara default, ketika men-download sebuah file, w3mir akan mencari URL di dalam file tersebut untuk proses download berikutnya. Dan, ini tidak hanya berlaku untuk file HTML, namun juga file PDF. Tentunya, proses seperti ini akan memakan waktu dan terkadang tidak selalu diinginkan. Apa yang Anda inginkan barangkali hanyalah men-download sebuah file dan tidak peduli dengan URL yang terdapat di dalamnya. Untuk kebutuhan seperti ini, Anda bisa memberikan opsi `-B` ketika men-download.

Harap diperhatikan bahwa opsi ini seharusnya tidak dikombinasikan dengan opsi `-r` yang akan menyebabkan w3mir men-download secara rekursif. Apabila Anda melakukannya, apa yang Anda download tidak akan benar-benar dapat digunakan sebagai mirror karena tidak ada URL yang dimanipulasi.

### ● Menangani javascript.

Bagaimana kalau HTML yang ingin di-download berisikan javascript? W3mir akan berusaha untuk tidak mengganggu script tersebut. Namun, w3mir tidak akan dapat benar-benar bekerja dengan javascript (dalam artian menjalankan javascript). Dengan demikian, hasil download mungkin tidak akan persis sama dengan yang tersimpan di server.

### ● Menangani CSS.

W3mir dapat bekerja dengan CSS, namun tidak mendukung sepenuhnya. Artinya, *style* yang didefinisikan dalam halaman HTML akan di-download apa adanya. Begitu pun

dengan file CSS eksternal. Yang tidak akan dilakukan oleh w3mir adalah menganalisis dan men-download apa yang terdapat dalam CSS (apabila ada).

### ● Menangani PDF.

W3mir, seperti sempat disebutkan sebelumnya, dapat pula bekerja dengan file PDF. W3mir dapat meng-*extract* URL yang tersimpan di dalam file PDF dan dapat men-download URL tersebut apabila diinginkan. Hanya, W3mir tidak dapat mengubah URL di dalam file PDF tersebut agar sesuai dengan kondisi mirror. Hal ini cukup repot untuk dilakukan dan bukan merupakan fungsi utama program.

### ● Menangani CGI.

W3mir tidak akan bisa men-download script CGI apa adanya karena script CGI tersebut akan dikerjakan dan hasilnya yang akan didapat oleh web server untuk dikirimkan ke client. Jadi, apa yang bisa Anda dapatkan hanyalah hasil pemrosesannya saja. Menggunakan W3Mir dengan protokol HTTP tidak akan membuat kita bisa mendapatkan semua file di web server.


### ● Bekerja dengan proxy server.

Untuk bekerja dengan proxy server, Anda dapat mempergunakan opsi `-P` ketika menjalankan w3mir. Berikut adalah contohnya:

```
$ w3mir -P proxy_server:8080
http://www.site.com/
```

Untuk proxy yang mewajibkan *username* dan *password*, Anda dapat menspesifikasikan *username* dan *password* tersebut di dalam file konfigurasi. Berikut ini adalah direktif yang digunakan dan contohnya:

```
HTTP-Proxy-user: proxy-username
HTTP-Proxy-passwd: proxy-password
```

Di Internet, terutama soal TI, kita bisa mendapatkan begitu banyak hal. Apabila kita tidak memiliki koneksi 24/7, kita tentu bisa mempergunakan program semacam w3mir dalam mengambil informasi (terutama secara berkala di kala koneksi sedang cepat). Hanya, masalah legal juga harus diperhatikan. Demikianlah pembahasan kita tentang w3mir. Selamat mencoba!   
Noprianto ([noprianto@infolinux.co.id](mailto:noprianto@infolinux.co.id))