

## Curl : l'Internet en ligne de commande.

Des logiciels pour récupérer des documents sur des serveurs via FTP ou HTTP, il en existe des dizaines, à commencer par votre navigateur Web favori. Mais vous est-il possible de lui commander un téléchargement depuis un script ?

Yannick Cadin

### Archivage en local de la page d'accueil d'un site Web

```
$ curl http://bsdnews.com/ > "Nouvelles BSD au `date +%d-%m-%Y`.htm"
```

Récupération via le protocole HTTP de la page d'entrée du site BSDnews et sauvegarde dans un fichier mentionnant la date du jour (format jj-mm-aaaa) dans le dossier courant.

### Liste d'un dossier sur un serveur FTP

```
$ curl ftp://ftp.apple.com/developer/Technical_Documentation/
```

Présentation détaillée des fichiers du dossier Technical\_Documentation du site FTP d'Apple consacré aux programmeurs.

### Affichage d'un document disponible sur un serveur FTP

```
$ curl -O ftp://sunsite.cnlab-switch.ch/mirror/yellowdog/iso/README
```

Chargement et présentation directe dans Terminal du fichier README présent dans le dossier iso d'un site offrant les images de CD de Linux YellowDog. Par défaut, curl envoie sur la sortie standard tous ses résultats et les documents qu'on lui demande de rapatrier.

### Téléchargements FTP multiples

```
$ curl -O "ftp://sunsite.cnlab-switch.ch/mirror/yellowdog/iso/yellowdog-3.0.1-sirius-20031118-install[1-3].iso"
```

Curl permet de spécifier des intervalles de caractères ou de valeurs (entre crochets), ici de 1 à 3, mais il faut alors protéger cette syntaxe pour éviter une tentative d'interprétation par le shell (ce qui explique la présence des guillemets). L'option -O (« output ») ordonne à curl d'enregistrer les documents récupérés dans des fichiers locaux portant les noms respectifs des fichiers d'origine plutôt que de les diriger vers la sortie standard (son comportement par défaut en l'absence de -O).

### Téléchargements groupés : autre formulation

```
$ curl -O "ftp://iso2.de.NetBSD.org/pub/NetBSD/iso/1.6.2/mac{68k,ppc}cd.iso"
```

Outre les intervalles entre crochets, curl permet aussi d'énumérer des textes (entre accolades) qui seront utilisés tour à tour pour produire une liste de noms distincts. Ici nous récupérons les fichiers mac68kcd.iso puis macppccd.iso, le tout en une seule commande curl. Notez les guillemets encore présents pour interdire au shell d'évaluer un texte comportant des accolades.

### Reprise de chargement au point d'interruption

```
$ curl -O -C - "ftp://iso2.de.NetBSD.org/pub/NetBSD/iso/1.6.2/mac{68k,ppc}cd.iso"
```

Le transfert a été coupé après le 562<sup>ème</sup> Mo ! Qu'à cela ne tienne, l'option -C (continue) de curl lève cette contrariété. Le tiret qui suit -C demande à curl de déterminer intelligemment le point de reprise.

### Envoi de document sur un serveur FTP

```
$ curl -T article.rtf -u redacteur:motdepasse ftp://redaction.diablotin.info/articles/svmmac/
```

Curl autorise aussi les « uploads » avec -T suivi du nom d'un fichier vers des serveurs Web (action PUT du protocole HTTP) ou FTP. En l'absence de l'option -u l'identifiant anonymous sera utilisé sur un site FTP.

Curl ne sait pas pour l'heure envoyer plusieurs fichiers en une seule invocation.

### Création de dossier PUIS envoi de fichier

```
$ curl -T fichier -u carole:secret ftp://serveur/nouveau/ -Q 'MKD nouveau'
```

En supplément, l'option -Q permet de soumettre des commandes directement au serveur FTP. Ce sont les commandes de « bas niveau » des serveurs FTP et peuvent varier plus ou moins d'une implémentation à l'autre. Une commande -Q qui échoue entraîne l'arrêt de curl.

### Invocation d'un script CGI sur un site Web sécurisé

```
$ curl -s 'https://gandi.net/whois?l=FR&domain=diablotin.com' | awk '/^owner-address:/ { print substr ($0, 16) }'
```

Curl sait aussi dialoguer via le protocole HTTPS en s'appuyant sur SSL. Ici un script CGI est sollicité (méthode GET du protocole HTTP) qui retourne la fiche descriptive d'un nom de domaine dont nous prélevons uniquement l'adresse du propriétaire à l'aide du programme awk.

### Soumission de formulaire Web

```
$ curl -A ptitmalin -s -d Amount=50 -d From=EUR -d To=USD http://www.xe.com/ucc/convert.cgi | sed -n 's/^.*<B>\([0-9.]*\) USD<.*$/\1/p'
```

Nous avons déjà rencontré ce cas avec netcat. C'est exactement la même chose mais exprimée cette fois avec la syntaxe de curl qui est plus spécialisée pour ce type de travail. L'affectation des différents champs du formulaire proposé par le site Xe.com se fait avec autant d'options -d NomDuChamp=Valeur (l'ordre n'est pas censé avoir la moindre importance). C'est la méthode POST du protocole HTTP qui est alors implicitement utilisée (par opposition à la méthode GET). L'option -s impose à curl de garder le Silence, il affiche sinon la progression du transfert pour le document rapatrié. L'option -A est nécessaire ici pour déjouer la méfiance du script convert.cgi de Xe.com qui interdit normalement son accès aux « robots » tels que curl (-A permet de changer le « User-agent » qui sert d'identité pour un programme de navigation Web). Finalement le résultat obtenu est filtré à l'aide du programme sed pour extraire juste le montant désiré.

### Seulement les en-têtes

```
$ curl -I http://formation.diablotin.biz/
```

Pour superviser le bon fonctionnement des sites Web vous pouvez avantageusement préférer récupérer l'en-tête (méthode HEAD du protocole HTTP) de chaque URL à contrôler. L'information retournée contient généralement (pour les pages statiques) la date de dernière actualisation.

### Juste une partie du document

```
$ curl -r 0-2000 http://www.diablotin.org/
```

Vous voulez faire votre propre moteur de recherche et vous optimisez en limitant aux 2000 premiers caractères de la page d'accueil de chaque site visité l'espace dans lequel vous trouverez sa description ainsi que sa liste de mots-clés dans les balises META correspondantes.

### Besoin d'explication ? Utilisez le dictionnaire

```
$ curl dict://dict.org/d:BSD:jargon
```

Dict fait partie des différents protocoles accessibles avec curl. Nous interrogeons ici l'ouvrage appelé jargon pour y trouver la définition (d:) du terme BSD sur le site dict.org.