

Les problèmes posés par les raccourcisseurs d'URL

Stéphane Bortzmeyer

<stephane+blog@bortzmeyer.org>

Première rédaction de cet article le 28 mars 2010

<https://www.bortzmeyer.org/probleme-raccourcisseurs-urls.html>

On voit partout des URL raccourcis, où le vrai URL a été remplacé par une version courte, précédée du nom du service de raccourcissement. `tinyurl.com` a été le premier mais, aujourd'hui, on voit plus souvent `goo.gl`, `bit.ly` ou `to` (certainement le plus court mais aussi le moins garanti <<https://www.bortzmeyer.org/domaines-d-un-seul-label.html>>). J'utilise le moins possible ces services.

Comment fonctionnent-ils? La première fois qu'on soumet un URL normal à un tel service (via un formulaire Web ou bien via l'API du service), un nombre aléatoire est créé et la correspondance entre ce nombre et le vrai URL est mise dans la base de données du service de raccourcissage. Ainsi, si je soumetts l'URL de cet article, <http://www.bortzmeyer.org/probleme-raccourcisseurs-urls.html> à `bit.ly`, j'obtiens un encodage du nombre (encodage choisi pour sa compacité), l'identificateur `bKulB8`, et un URL formé à partir de cet identificateur, <http://bit.ly/bKulB8>. Si je suis ensuite cet URL raccourci, je me connecte au service raccourcisseur qui utilisera la redirection standard de HTTP (RFC 7231¹, section 6.4.2, au sujet du code 301) pour m'envoyer au vrai URL. On le voit bien avec `wget` :

```
% wget http://bit.ly/bKulB8
--2010-03-28 22:08:12-- http://bit.ly/bKulB8

Resolving bit.ly... 128.121.234.46, 128.121.254.129, 128.121.254.201, ...
Connecting to bit.ly|128.121.234.46|:80... connected.
HTTP request sent, awaiting response... 301 Moved
Location: http://www.bortzmeyer.org/probleme-raccourcisseurs-urls.html [following]

--2010-03-28 22:08:13-- http://www.bortzmeyer.org/probleme-raccourcisseurs-urls.html
Resolving www.bortzmeyer.org... 208.75.84.80, 2001:4b98:41::d946:bee8:232
Connecting to www.bortzmeyer.org|208.75.84.80|:80... connected.
HTTP request sent, awaiting response... 200 OK
Length: 10011 (9.8K) [text/html]
Saving to: ...
```

1. Pour voir le RFC de numéro NNN, <https://www.ietf.org/rfc/rfcNNN.txt>, par exemple <https://www.ietf.org/rfc/rfc7231.txt>

Il existe quelques bonnes raisons d'utiliser ces services : la principale concerne les services qui limitent sérieusement la taille des messages, le plus connu étant Twitter. Beaucoup d'URL dépassant à eux seuls la fameuse limite des 140 caractères, l'usage d'un raccourcisseur est crucial et, de toute façon, il est automatique. Même si vous essayez de publier votre propre URL, Twitter le raccourcira automatiquement (aujourd'hui avec `bit.ly`).

Les autres raisons sont moins bonnes : par exemple, si le nom de l'URL donne déjà une indication du contenu, on peut utiliser le raccourcisseur pour maintenir la surprise plus longtemps. Autre raison acceptable, certains sites Web sont mal conçus et les URL qu'ils présentent lors de la navigation sont anormalement longs (c'est par exemple le cas de Légifrance). La bonne solution, pour ce dernier cas, serait que les gérants de ces sites utilisent de beaux URL `<https://www.bortzmeyer.org/beaux-urls.html>`.

Mais, en dehors de ces quelques points ponctuels, les raccourcisseurs n'ont que des défauts :

- Ils permettent à leurs promoteurs de surveiller la navigation des internautes, puisque tout accès au site va nécessiter un passage par le raccourcisseur. Ils sont donc une menace pour la vie privée. (Cette menace peut être bien illustrée par les statistiques qu'offre le raccourcisseur, par exemple, pour `bit.ly`, en ajoutant un + à la fin de l'URL. Si vous allez en `<http://bit.ly/info/bKulB8+>`, vous verrez combien de gens ont utilisé le lien court, depuis quel pays, etc.)
- Ils n'offrent aucun indice sur leur destination, ce qui ne permet plus d'analyser avant de cliquer l'endroit où on va arriver et son sérieux. (Cela peut également avoir des conséquences pour la sécurité `<http://blog.trendmicro.com/shortened-urls-in-im-apps-lead-to-a-worm/>`. Voir une bonne application de ce problème avec le méchant raccourcisseur `<http://spareclockcycles.org/2010/12/19/d0z-me-the-evil-url-shortener/>`.)
- Ils menacent la permanence des liens, puisque ils ajoutent un intermédiaire de plus, dont rien n'indique s'il va garder sa base longtemps. Il ne faut donc jamais mettre un URL raccourci dans un document voué à rester stable.
- Comme les URL raccourcis sont composés uniquement de chiffres aléatoires, on ne peut pas les mémoriser ou les deviner.
- Ils interdisent de savoir si on a déjà vu ou non tel ou tel article, puisque chaque service de raccourcissage va créer son propre URL raccourci.

Heureusement, il existe, grâce à `fil@rezo.net`, un service de rallongement des URL `<http://long.rezo.net/>`, qui permet, en connaissant un URL raccourci, de retrouver le vrai (voir l'article « Allongez vos URLs `<http://zzz.rezo.net/Allongez-vos-URLs.html>` »). Vous pouvez même utiliser ce service automatiquement.

Un service équivalent, plus ancien, est `longurl.org`, qui dispose d'une API `<http://longurl.org/api>`, et de nombreux outils `<http://longurl.org/tools>`. `longurl.org` ne semble plus en état de marche aujourd'hui mais un autre outil, plus récent, est `longURLplease.com` que je ne trouve pas pratique à essayer : il n'a pas de formulaire Web, il faut passer par l'API. Il y a aussi `<http://untiny.me/>`, `<http://www.expandmyurl.com/>` et `<http://ghill.customer.netspace.net.au/embiggen/>` (ce dernier ne semble marcher qu'avec TinyURL et son interface Web ne marche pas pour moi.)

Pour le problème de sécurité mentionné plus haut (le gérant du raccourcisseur peut surveiller la navigation des internautes) et uniquement pour ce problème, une solution est d'installer une plate-forme de raccourcissement d'URL chez soi, comme YOURLS `<http://www.framasoft.net/article5004.html>`. Un tel système laisse intacts tous les autres problèmes...

Autres articles sur la question : « Le contrôle des redirections `<http://www.framablog.org/index.php/post/2010/10/31/le-controle-des-redirections>` », qui inclut une traduction de l'excellent « *Is the Web heading toward redirect hell?* » `<http://royal.pingdom.com/2010/09/`

22/is-the-web-heading-toward-redirect-hell/> », « **Allongez votre URL** <<https://signal.eu.org/blog/2010/01/08/allongez-votre-url/>> » de Pierre Beyssac, « **Obtenir l'URL réelle des URL courtes** <<http://blogzinet.free.fr/blog/index.php?post/2010/05/09/Obtenir-l-URL-reelle>> et bien sûr l'article historique « *"Terror of tinyurl"* <<http://www.wordyard.com/2007/10/05/terror-of-tinyurl/>> » de Scott Rosenberg, qui défend la valeur des URL compréhensibles.

Merci à Vincent Hinderer pour ses suggestions.