

# Les beaux URL

Stéphane Bortzmeyer

<stephane+blog@bortzmeyer.org>

Première rédaction de cet article le 10 janvier 2008. Dernière mise à jour le 14 janvier 2008

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

---

Beaucoup de sites Web affichent, dès qu'on dépasse la page d'accueil, et même parfois dès la page d'accueil, des URL incompréhensibles, très longs et bourrés de paramètres dont l'utilisateur ne connaît pas la signification. C'est déplorable, à la fois pour des raisons esthétiques et pour des raisons pratiques. Pourquoi est-ce regrettable ? Comment éviter cela ?

Prenons quelques exemples affreux. L'UNESCO proclame 2008 l'« Année Internationale des Langues » et la page Web de présentation de l'activité est à l'URL <[http://portal.unesco.org/culture/en/ev.php-URL\\_ID=35559&URL\\_DO=DO\\_TOPIC&URL\\_SECTION=201.html](http://portal.unesco.org/culture/en/ev.php-URL_ID=35559&URL_DO=DO_TOPIC&URL_SECTION=201.html)>. Pourquoi ne pas avoir utilisé le plus court et plus parlant <http://www.unesco.org/iyl?>. Autre exemple, si on se promène sur le site de Dell à la recherche d'informations sur le Latitude 430 <<https://www.bortzmeyer.org/dell-latitude-d430-linux.html>>, on arrive en <[http://premierconfigure.euro.dell.com/dellstore/config.aspx?cs=RC1077983&customer\\_id=RC1077983&oc=LD4301&~tgt=global\\_cfg&c=FR&l=fr&s=PAD](http://premierconfigure.euro.dell.com/dellstore/config.aspx?cs=RC1077983&customer_id=RC1077983&oc=LD4301&~tgt=global_cfg&c=FR&l=fr&s=PAD)>. Pourquoi ne pas avoir un URL comme <http://www.dell.com/computers/latitude>. Un dernier exemple, alors que la loi est censée être facilement accessible à tous, il n'existe pas de moyen de fabriquer facilement l'URL pour, mettons le « Décret n°2006-1386 du 15 novembre 2006 fixant les conditions d'application de l'interdiction de fumer dans les lieux affectés à un usage collectif ». Sur Légifrance, on arrive à un URL de <<http://www.legifrance.gouv.fr/affichTexte.do?cidTexte=JORFTEXT000000818309&dateTexte=20080124&fastPos=5&fastReqId=621185253&oldAction=rechTexte>>. Pourquoi pas <http://www.legifrance.gouv.fr/decrets/2006/11/15/1386?> Manifestement, le webmestre de Légifrance a lui-même des problèmes avec de tels URL puisque, lorsqu'on regarde l'article R3512-1 du Code de la Santé Publique, et qu'on voit un beau lien vers le décret, ce lien ne fonctionne pas « Aucun document ne correspond à votre demande » !

C'est très laid. De tels URL heurtent mon sens esthétique. Pour se remonter le moral, voici quelques exemples de beaux URL :

- <[http://fr.wikipedia.org/wiki/Valeria\\_Bruni-Tedeschi](http://fr.wikipedia.org/wiki/Valeria_Bruni-Tedeschi)>. À part la mention "wiki" qui semble tout à fait inutile, les URL de Wikipédia sont courts et le sujet est connu à l'avance.
- <<http://embruns.net/carnet/actus-et-opinions/miss-france-2007.html>>. Comme souvent sur les blogs, un URL parlant, qui résume le sujet.

- <http://www.rueducommerce.fr/Ordinateurs/Ordinateur-Portable/Ordinateur-Portable-LENOVO/419103-Ordinateur-Portable-Lenovo-3000-N200-Intel-Celeron-M-530-1-73-GHz-htm>. Les sites commerciaux peuvent avoir des URL corrects, eux aussi, même si celui-ci est probablement trop détaillé, incluant même la vitesse du processeur..
- [http://www.humanite.fr/2007-06-14\\_Societe\\_En-Chine-Google-censure-a-tour-de-cl](http://www.humanite.fr/2007-06-14_Societe_En-Chine-Google-censure-a-tour-de-cl). L'URL est parlant : il inclut la date (ce qui est logique pour un quotidien) et un résumé du titre (comme le font souvent les moteurs de blog).
- <http://del.icio.us/bortzmeyer/fran%C3%A7ais>. Les URL de del.icio.us sont excellents (ici, le nom de l'utilisateur et le "tag" choisi) mais un problème technique a rendu celui-ci peu lisible. Originellement (RFC 1738<sup>1</sup>), les URL devaient s'exprimer uniquement en ASCII. Les IRI du RFC 3987 ont supprimé cette restriction mais ils ne sont pas encore très répandus (par exemple, mon validateur RelaxNG ne me permet pas d'en mettre dans ce blog car, dit-il, "*Relax-NG validity error : Type anyURI doesn't allow value 'http ://del.icio.us/bortzmeyer/français'*"). Essayez `http://del.icio.us/bortzmeyer/français` pour voir si votre navigateur gère les IRI.
- <http://www.figoblog.org/node/1903>. L'URL (pris sur le site Figoblog <http://www.figoblog.org/>) est court et propre mais pas décodable (pourquoi 1903? Ce n'est qu'un identificateur interne.) Mais c'est cohérent avec le discours de son auteure, qui estime qu'un identifiant stable ne doit pas avoir de sémantique (car un changement de sémantique pourrait changer l'URL).

Outre l'aspect esthétique, les URL laids posent aussi des problèmes pratiques (également bien détaillés dans un excellent article <http://www.useit.com/alertbox/990321.html> de Jakob Nielsen). Étant longs, ils sont difficiles à copier-coller, par exemple dans un courrier. Révélant le fonctionnement interne du site Web (noms des paramètres, type de langage utilisé comme PHP pour le site de l'UNESCO), ils ne sont pas stables. Si le site passe à un autre CMS, ces URL changeront probablement, en contradiction avec les principes posés dans l'article "*Cool URIs don't change*" <http://www.w3.org/Provider/Style/URI>.

Surtout, ces URL sont « indécodables ». Avec la meilleure volonté du monde, on ne peut pas les comprendre, on ne peut pas les mémoriser et on ne peut pas les bricoler de manière créative, par exemple en changeant la valeur d'un paramètre.

De tels URL sont parfois volontaires, par exemple pour décourager la création de « liens profonds » (liens hypertexte pointant vers une autre page que la page d'accueil). Pourquoi diable voudrait-on décourager les liens profonds? Typiquement pour empêcher le lecteur d'accéder directement à l'information qui l'intéresse et pour l'obliger à suivre le parcours officiel, partant de la page d'accueil.

De tels URL très longs sont parfois volontaires parce qu'il y a beaucoup d'informations à indiquer. C'est le cas des URL de Google Maps où il y a beaucoup à encoder (notamment longitude et latitude). Par exemple, dans `http://www.google.com/maps?f=q&hl=en&geocode=&time=&date=&ttype=&q=san+cristobal+de+las+casas,+chiapas,+mexico&sll=16.75655,-93.129044&sspn=0.425394,0.579529&ie=UTF8&ll=16.738551,-92.633286&spn=0.053179,0.072441&t=h&z=14&iwloc=addr&om=1`, les chiffres dans le paramètre `ll` indiquent la latitude et la longitude. Mais de tels URL ne sont pas gratuitement compliqués, ils stockent de l'information et ils sont décodables et modifiables. Notamment, je peux générer un tel URL facilement, sans avoir besoin de Google Maps, ce qui permet des applications géographiques intéressantes de type "*mashup*". Par exemple, comme le service Vélib' <https://www.bortzmeyer.org/velib-rest.html> donne accès à la latitude et la longitude de chaque station (dans le fichier <http://www.velib.paris.fr/service/carto>), on peut facilement faire un programme qui crée un URL Google Maps pour chaque station, comme c'est le cas sur des sites comme Roulib <http://www.roulib.fr/> ou bien <http://v.mat.cc/>. D'ailleurs, il existe même un langage pour décrire ces gabarits d'URL, normalisé dans le RFC 6570.

---

1. Pour voir le RFC de numéro NNN, <https://www.ietf.org/rfc/rfcNNN.txt>, par exemple <https://www.ietf.org/rfc/rfc1738.txt>

Mais, la plupart du temps, ces URL sont complexes car le webmestre a installé un logiciel de gestion du site Web (qu'il n'a pas toujours choisi) et que ce sont les URL par défaut du logiciel et que le webmestre ne peut pas ou ne sait pas les modifier (ou, pire, n'a pas compris à quoi servaient les URL <<https://www.bortzmeyer.org/identificateur-vs-moteur-de-recherche.html>>).

A contrario, certains logiciels, comme la plupart des moteurs de blogs, fabriquent des URL clairs et simples, et faciles à décoder. Ils ne sont pas forcément stables sur le long terme (les professionnels de la stabilité suggèrent en général de ne pas mettre trop d'information dans l'URL car un changement de cette information changerait l'identificateur) mais ils sont bien agréables à manipuler.

Alors, que peut-on faire pour ne pas avoir de tels URL? Si le webmestre est impossible à faire évoluer sur ce sujet (souvent, il ne lit même pas ses messages <<https://www.bortzmeyer.org/personne-ne-s-est-plaint.html>>), il existe des mécanismes de contournement comme tinyurl. Ces systèmes créent une correspondance entre un URL et un identificateur à eux, et fournissent un service de redirection automatique. C'est ainsi que <<http://tinyurl.com/2u2ewd>> va au même endroit que l'affreux URL de l'UNESCO cité plus haut. Mais ces mécanismes ont leurs propres inconvénients <<https://www.bortzmeyer.org/probleme-raccourcisseurs-urls.html>>, notamment le fait que l'URL soit encore moins décodable qu'avant (on ne voit même pas le nom de domaine, donc on ne peut pas savoir qui est responsable).

L'idéal est donc quand le webmestre fait bien son travail. Pour l'aider, voyons quelques techniques utiles. Notons d'avord qu'il n'existe aucune raison technique valable d'utiliser des URL affreux. On voit parfois certains débutants croire que des URL incluant un point d'interrogation sont nécessaires si la recherche de l'information est dynamique (par exemple <http://www.maboutique.fr/catalogue?objet=642342>). Mais c'est faux, l'URL est indépendant de la méthode utilisée pour générer la page. On peut avoir un URL qui va faire une recherche dynamique dans une base de données, sans point d'interrogation dedans. (On notera que, soucieux de prouver chaque jour qu'on peut breveter n'importe quoi, l'office états-unien des brevets a enregistré un brevet n° 7,287,042 <<http://patft.uspto.gov/netacgi/nph-Parser?Sect1=PTO1&Sect2=HITOFF&d=PALL&p=1&u=%2Fnetahhtml%2FFPTO%2Fsrchnum.htm&r=1&f=G&l=50&s1=7,287,042.PN.&OS=PN/7,287,042&RS=PN/7,287,042>> qui réserve cette technique à Amazon, en dépit de nombreux précédents d'utilisation, que l'incompétence ou la malhonnêteté de l'office des brevets ne lui a pas permis de détecter.)

Notons d'abord que beaucoup de CMS ou de logiciels équivalents permettent de faire facilement des beaux URL. Si ce service n'est pas activé par défaut, il est souvent bien documenté. Par exemple, le CMS SPIP a une documentation détaillée « Utiliser des URLs personnalisées <[http://www.spip.net/fr\\_article765.html](http://www.spip.net/fr_article765.html)> ».

Sinon, la solution générale, si on n'est pas l'auteur du CMS utilisé ou bien qu'on n'a pas envie de se plonger dans le code source est, si on utilise Apache, de se servir de l'excellent (mais complexe) module `mod_rewrite`.