

<b>Domaine :</b>	Recherche	<b>Référencement</b>
<b>Niveau :</b>	Pour tous	<b>Avancé</b>

*WordPress est un CMS prisé par les référenceurs et les utilisateurs pour sa simplicité d'utilisation et la facilité à l'optimiser pour le référencement naturel. Malheureusement, celui-ci a une lacune importante : le contenu dupliqué. Voici comment le limiter ou même totalement l'éradiquer sur votre site en configurant au mieux votre plate-forme...*

Dans cet article, nous allons voir comment réduire au minimum le phénomène de duplicate content, à la fois pour l'internaute et le moteur de recherche.

*PS : nous partirons du principe que vous utilisez le plugin WordPress SEO, l'un des meilleurs et des plus complets plugins de référencement naturel pour WordPress : <http://wordpress.org/extend/plugins/wordpress-seo/>*

## C'est quoi le duplicate content ?

Avant tout chose, il est bon de rappeler quelques bases. Le contenu dupliqué est le fait de trouver un contenu identique sur plusieurs adresses web différentes. Celui-ci peut-être :

- interne :
  - pageA.html
  - pageB.html
- externe :
  - mon-site.com/pageA.html
  - site-concurrent.com/pageB.html
- partiel ou complet (le contenu a pu être dupliqué intégralement ou partiellement)

### En externe

Il faut bien différencier le contenu dupliqué en interne de celui qui ne l'est pas. Situé sur un autre nom de domaine que le vôtre, Google doit déterminer quelle est la copie originale. Pour cela, il peut utiliser les dates de publication, la popularité et la "crédibilité" de chaque site. Malgré les différents signaux que peut utiliser le moteur de recherche, il est fréquent pour lui de se tromper.

Si le moteur de recherche estime que votre site est celui qui copie, il le pénalisera. En d'autres termes, le contenu dupliqué externe est le plus nuisible d'entre tous.

C'est aussi le plus dur à combattre : un simple copier/coller - plus ou moins automatisé - d'une personne peu scrupuleuse permet de reproduire facilement toutes vos publications.

### En interne

La duplication interne est moins dangereuse car elle ne provoque pas de pénalité directe sur votre référencement naturel. Celle-ci est plus insidieuse. Un contenu identique sur plusieurs pages de votre site va en effet avoir plusieurs effets indésirables :

- Cela ralentit l'indexation des contenus : pourquoi continuer à indexer si vos contenus sont identiques ?
- Cela force Google à choisir quelle page positionner parmi les différentes URL disponibles (et il ne choisit pas forcément la plus pertinente).
- Cela dilue la popularité du site sur des pages similaires et inutiles.

Vous n'aurez jamais de pénalité en tant que tel pour une duplication de contenu interne, mais cela peut quand même nuire à votre référencement.

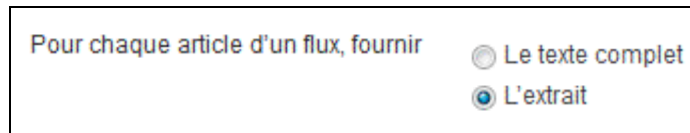
## WordPress duplique le contenu en externe

Comme expliqué précédemment, il est difficile de lutter contre la duplication de contenu sur d'autres sites. Rien ne vous protégera entièrement du simple copier/coller.

Cependant, une forte partie du contenu dupliqué en externe est réalisé avec des logiciels automatisés, qui vont scanner votre site et vos articles pour les récupérer. Et le premier point d'entrée de ces robots est votre flux RSS.

Par défaut, WordPress affiche l'intégralité de chaque article dans votre flux d'actualité, à l'adresse suivante : <monsite.com/feed/>

En changeant une simple option, le CMS n'affichera plus qu'un extrait de contenu. Cela réduira considérablement les risques de contenus dupliqués, tout en incitant vos lecteurs à se rendre sur votre site plutôt qu'à utiliser un lecteur de flux RSS. Rendez-vous dans le menu "**Réglages**" => "**Lecture**", puis cocher l'extrait pour l'option "*Pour chaque article d'un flux, fournir*".



La deuxième méthode consiste à installer un plugin appelé **Read More Copy Link** : <http://wordpress.org/extend/plugins/read-more-copy-link/screenshots/>

Celui-ci va permettre :

- d'ajouter un lien "*Lire la suite*" dans tous les texte copiés avec Ctrl-C.
- de réduire la quantité de texte copié (en fonction du paramétrage que vous souhaitez).

## Le contenu dupliqué en Interne

Avant d'entrer dans le détail de chaque type de page, il faut savoir une chose sur WordPress : sans paramétrage spécifique et avec un thème par défaut, un article publié sera dupliqué à plusieurs niveaux. On retrouvera ainsi l'intégralité du contenu sur :

- La page de l'article lui-même.
- La page d'accueil.
- Les pages de catégorie.
- Les pages de mots clés.
- La page d'auteur.
- La page d'archive par date.
- Les pages de Custom Taxonomies (nous en parlerons en fin d'article).
- Les sous-pages de chacune des pages citées précédemment.

Il faut donc trouver un moyen de présenter soit un contenu différent, soit un contenu tronqué sur chacune de ces pages.

### Les articles et l'accueil du site

La première étape cruciale est de ne pas dupliquer l'intégralité du contenu de chaque article. Pour faire cela, il faut utiliser une simple option disponible par défaut lors de la rédaction d'un article : le bouton "<!--more-->". Celui-ci se situe parmi les autres boutons de mise en page :

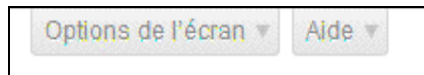


Quand vous l'insérez dans une publication, tout le contenu situé avant la balise sera visible comme avant sur toutes les pages du site, mais celui situé après ne sera visible que dans l'article lui-même. Les autres pages du site vont donc afficher un lien "*Lire la suite*" pour accéder à l'intégralité du contenu. On ne duplique donc plus que le début de chaque article et non plus son intégralité.

### Les catégories

Une fois ce premier problème réglé, nous avons toujours un contenu identique sur l'accueil et les autres types de pages. Nous allons donc nous attaquer tout d'abord aux catégories.

Là aussi, on peut réussir facilement à éradiquer le contenu dupliqué. Pour cela, il faut commencer par rédiger une description courte de chacun de vos articles grâce à ce qu'on appelle l'**excerpt** (traduit en français par "*Extrait*"). Par défaut, ce bloc est caché lors de la rédaction d'un contenu. Pour l'afficher, un menu d'option se trouve en haut à droite.



Une fois cette description renseignée, il faudra modifier certains fichiers de votre thème, situé dans le répertoire [monsite.com/wp-content/themes/nom-du-theme/](http://monsite.com/wp-content/themes/nom-du-theme/)

Il faut en effet remplacer toute référence à `the_content()` par `the_excerpt()`. En fonction de votre thème, il est possible de trouver les références incriminées dans les fichiers suivants :

- `category-nom.php`
- `category-id.php`
- `category.php`
- `archive.php`, si aucune présence des fichiers `category.php`
- `index.php`, si aucune présence des fichiers `category.php` et `archive.php`. Faites alors attention à placer ce code avec des conditions du type : `if ( is_category() ) the_excerpt();`

Attention, `the_content()` peut parfois avoir des paramètres situés entre les parenthèses mais il est inutile de les recopier dans `the_excerpt()`.

Malheureusement, c'est encore insuffisant pour avoir des catégories dignes de ce nom, car les pages de catégories possèdent encore d'autres lacunes.

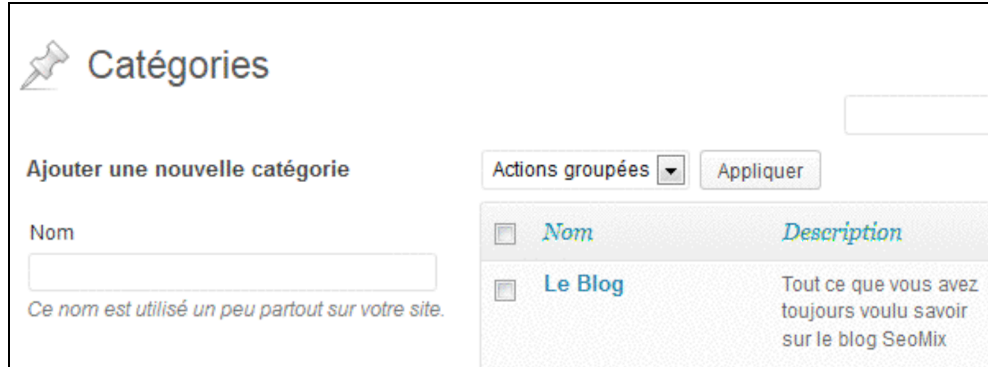
La première d'entre elle est que WordPress va tester la dernière partie de l'URL de la catégorie pour savoir si elle existe ou non. Ainsi, si vous tapez [monsite.com/category/nom-de-la-categorie-36/](http://monsite.com/category/nom-de-la-categorie-36/) et [monsite.com/category/nom-de-la-categorie-36/nom-de-la-categorie-36/](http://monsite.com/category/nom-de-la-categorie-36/nom-de-la-categorie-36/), les deux URL vont renvoyer le même contenu. Pour éviter cela, il existe plusieurs possibilités :

- supprimer le `/category/` par défaut des URL de catégorie. Pour faire cela, il existe deux méthodes différentes :
  - une option dans le plugin **WordPress SEO** via le menu **SEO => Permalien**
  - un plugin qui ne fait que cela : **WP No Category Base** (<http://wordpress.org/extend/plugins/wp-no-category-base/>)
- utiliser une solution donnée par *Julio Potier* de **BoiteAWeb**, qui consiste à ajouter un code dans le fichier `functions.php` de votre thème WordPress : <http://www.geekpress.fr/wordpress/astuce/erreur-404-pagination-1422/>

La seconde lacune est d'avoir le même contenu dans toutes les catégories associées à votre article. Si une publication est associée à la catégorie A et B, le contenu sera

identique sur les deux pages de catégorie. Le mieux et le plus simple pour éviter cela est de bien concevoir la structure de votre site et de vos contenus. Chaque catégorie doit être logique et unique, le but étant de n'avoir qu'une seule catégorie associée à chaque article.

La troisième lacune est qu'une catégorie devient alors une succession d'extraits d'articles, sans avoir de réel contenu unique et pertinent supplémentaire. On peut agir de plusieurs façons sur cela : on peut tout d'abord remplir la description de la catégorie dans l'administration de WordPress.



Il suffit ensuite de l'afficher avec le code suivant, toujours dans les mêmes fichiers que précédemment.

```
<?php echo category_description(); ?>
```

On peut également opter pour le fait d'avoir des affichages complètement différents par catégorie. Comme on l'a vu dans la liste des fichiers, on peut associer une mise en page et des contenus en fonction d'un ID ou d'un nom de catégorie. Par exemple, je peux très bien avoir :

- `category-36.php`, un fichier de template unique pour la catégorie 36.
- `category-38.php`, un autre template pour la catégorie 38.
- ...
- `category.php`, pour l'affichage par défaut de toutes les autres catégories.

Libre à vous de modifier chaque template en fonction de vos besoins et objectifs.

### Les tags (ou mots clés)

Les étapes suivantes consistent à améliorer les pages de mots clés. Là aussi, nous allons tout d'abord chercher à ne pas avoir le même contenu. Pour gagner du temps, il est recommandé d'appliquer la même méthodologie que pour les catégories, à savoir :

- Remplissez la description de vos mots clés dans l'administration de WordPress, puis affichez-les avec le code `<?php echo tag_description(); ?>`
- Utiliser le contenu tronqué avec l'aide de `the_excerpt()` au lieu de `the_content()`

Toutes ces modifications doivent être réalisées dans les fichiers suivants de template de votre thème :

- `tag-nom.php`
- `tag-id.php`
- `tag.php`
- `archive.php`, si aucune présence des fichiers `tag.php`
- `index.php`, si aucune présence des fichiers `tag.php` et `archive.php`. De même, faites attention à placer votre code avec des conditions du type : `if ( is_tag() ) the_excerpt();`

On peut éventuellement aller plus loin si l'on veut vraiment n'avoir aucun texte dupliqué. Rappelez-vous que `the_excerpt()` reprend l'extrait de votre article, mais le contenu sera donc identique sur la page de catégorie et celles des mots clés. Là aussi, PHP nous donne la possibilité de prendre le contenu intégral d'un article en le tronquant après un nombre défini de caractères (260 dans l'exemples ci-dessous). Voici tout d'abord le code à placer dans le fichier `functions.php` de votre thème :

```
function seomix_description() {  
    global $post;  
    $contenu = strip_tags($post->post_content);  
    $contenu = strip_shortcodes($contenu);  
    $contenu = str_replace(array("\n", "\r", "\t"), ' ',  
$contenu);  
    $contenu = '<p>'.substr($contenu, 0, 260).'...</p>';  
    echo $contenu;  
}
```

Placer ensuite le code suivant en remplacement de `the_content()` ou de `the_excerpt()` :

```
<?php if (function_exists('seomix_description')) {  
seomix_description();}?>
```

Mais attention, même en faisant les étapes précédentes les mots clés peuvent se révéler être un véritable fléau pour le contenu dupliqué, pour la simple et bonne raison que bien souvent les utilisateurs "adorent" ajouter un grand nombre de mots clés plus ou moins pertinents à leurs articles.

Un mot clé pertinent doit donc respecter plusieurs règles :

- Etre compréhensible en dehors du contexte : le mot clé "*Conseil*" ne veut rien dire alors que le mot clé "*Conseils allaitement*" a beaucoup plus de sens.
- Il ne doit pas être identique à un nom de catégorie (car cela sera un pur doublon).
- Il ne doit pas non plus être identique au titre d'un article (car dans ce cas de figure vous compliquez la tâche de Google pour savoir quel page il doit afficher dans les résultats du moteur de recherche).
- Il ne doit pas être similaire à un autre mot clé : par exemple il faudrait éviter de voir coexister des mots clés trop proches, comme avec des variantes singulier/pluriel ou des fautes d'orthographe. Par exemple, évitez les mots clés "*WordPres*" et "*Word Press*".
- Ils ne doivent pas être trop nombreux, car cela implique forcément d'avoir le même contenu sur plusieurs pages de mots clés. Si votre publication possède 10 mots clés associés, vous aurez donc dix fois le même contenu dupliqué (même si cette duplication est partielle).

Rappelez-vous toujours que Google ne juge pas un site sur le nombre de contenus, mais sur la popularité et la pertinence de ceux-ci : plus il y aura des contenus dupliqués, plus il est probable que chacun d'entre eux soit moins pertinent et moins populaire.

### Les dates

WordPress va également créer par défaut des archives par date. Elles fonctionnent comme les catégories et les mots clés dans le sens où elles listent vos articles, provoquant de la même manière un contenu dupliqué supplémentaire.

Pire encore, l'aspect sémantique est ignoré car les dates de publication n'impliquent pas forcément un lien logique entre chaque article. Par exemple, l'article A publié en Février peut parler par exemple d'écologie et l'article B publié le même mois peut parler du dernier concert des Spice Girls. Dans plus de 9 cas sur dix, votre site Internet sera dans ce cas de figure.

La solution la plus adéquate est donc d'éradiquer purement et simplement ce type d'archives. Pour cela, il faut effectuer plusieurs actions :

- Dans le menu **"Apparence" => "Widgets"**, vérifiez qu'aucun contenu et widget ne fait référence à ces archives par date.
- Faites de même dans le menu **"Apparence" => "Menus"**.
- Dans presque tout les fichiers de votre thème, vérifiez qu'aucun lien n'est fait vers ce type d'archive.
- Dans le plugin WordPress SEO, rendez-vous dans le menu **"SEO" => "Titres & Métas" => "Autre" => "Archives par date"**, puis cochez *"désactiver les archives par date"*.

Attention, si ces archives étaient actives et indexées par Google, il vous faudra mettre en place des redirections 301 de celles-ci vers l'accueil de votre site, soit avec un fichier htaccess, soit avec l'excellent plugin Redirection : <http://wordpress.org/extend/plugins/redirection/>

### Les auteurs

Les deux derniers types de contenus vont concerner un nombre plus restreint de sites Internet mais peuvent également nuire au référencement naturel. Le premier type est la page d'auteur. Sur une installation standard, une page de ce type contient la liste des articles publiés par l'utilisateur concerné.

Si vous êtes dans le cas de figure d'un site avec un seul auteur, et que vous ne souhaitez pas spécialement vous positionner sur le nom, prénom ou pseudo de celui-ci, cette page n'a aucun intérêt, exactement comme les archives par date. Il faudra donc supprimer tout lien pointant vers ce type de pages. Procédez de la même manière que pour une archive par date :

- Dans le menu **"Apparence" => "Widgets"**, vérifiez qu'aucun contenu ne fait référence aux auteurs
- Faites également de même dans le menu **"Apparence" => "Menus"**.
- Dans presque tout les fichiers de votre thème, vérifiez qu'aucun lien n'est fait vers les pages auteurs.

A l'inverse, si vous avez plusieurs auteurs ou que vous voulez vous positionner sur le nom de cette personne, il va falloir modifier cette page pour lui apporter un réel contenu. En fonction de votre thématique, les possibilités sont plus ou moins nombreuses mais voici quelques suggestions que vous pouvez mettre en place :

#### - Ajouter les 5 derniers commentaires de l'auteur

Dans le fichier `author.php` de votre thème, ajoutez ce code en haut de la page

```
<?php $curauth = (isset($_GET['author_name'])) ? get_user_by('slug',  
$author_name) : get_userdata(intval($author));  
$myid = $curauth->ID; query_posts("author=$myid&posts_per_page=-  
1");?>
```

Copiez ensuite à l'endroit désiré le code suivant :

```
<?php  
$comments = get_comments("author_email=$curauth-  
>user_email&number=5");  
foreach($comments as $comment) :  
    $theid = $comment->comment_post_ID;  
    echo '<blockquote><p>'. $comment-  
>comment_content. '</p></blockquote>  
    <p>Le '.get_comment_date( 'd F Y', $comment->comment_ID ).'  
dans <strong><a href="'.get_permalink($theid).'"
```

```
title="'.get_the_title($theid).'">'.get_the_title($theid).'    endforeach;  
?>
```

### - Ajouter une image de grande taille pour la photo

Cette étape est relativement simple. Copiez le code ci-dessous dans le fichier `author.php` de votre thème WordPress :

```
<?php echo get_avatar(get_the_author_meta('ID'), '150'); ?>
```

### - Ajouter une description plus étoffée de l'auteur

Pour réellement étoffer le contenu, on peut afficher ici une description plus longue de l'auteur. Pour y parvenir, il est nécessaire de suivre trois étapes.

1. Dans le fichier `functions.php` de votre thème copiez le code suivant :

```
//***** Infos supplémentaires sur les auteurs  
add_action( 'show_user_profile', 'extra_user_profile_fields' );  
add_action( 'edit_user_profile', 'extra_user_profile_fields' );  
  
function extra_user_profile_fields($user) { ?>  
<h3>Informations complémentaires sur le profil</h3>  
<table class="form-table"><tr><th><label for="address">Description  
détaillée de l'utilisateur</label></th><td>  
<textarea name="shortdesc" id="shortdesc" cols="8" rows="3"><?php  
echo esc_textarea(get_the_author_meta('shortdesc', $user->ID) );  
?></textarea><br />  
</td></tr></table>  
<?php }  
add_action( 'personal_options_update',  
'save_extra_user_profile_fields' );  
add_action( 'edit_user_profile_update',  
'save_extra_user_profile_fields' );  
  
function save_extra_user_profile_fields( $user_id ) {  
    if ( !current_user_can( 'edit_user', $user_id ) ) {  
        return false;}  
    update_usermeta( $user_id, 'shortdesc', $_POST['shortdesc'] );  
}
```

2. Dans l'administration de WordPress, vous aurez accès à un nouveau champ de description pour chaque profil utilisateur que vous devrez remplir.

3. Copiez pour finir le code suivant dans le fichier `author.php` de votre thème :

```
<?php the_author_meta('shortdesc');?>
```

### **Les custom taxonomies**

Ce dernier cas de figure concerne une proportion encore plus réduite de sites Internet. Avant tout, il faut expliquer de quoi il s'agit. Dans WordPress, une taxonomie est un moyen de classer les contenus. Chacun de vos articles possède donc 4 taxonomies par défaut :

- Les catégories.
- Les mots clés.
- L'auteur.
- La date de publication.

Il s'agit de 4 moyens différents utilisés pour classer et afficher vos différents contenus. Certains plugins ou thèmes peuvent créer leurs propres taxonomies. Vous les verrez théoriquement apparaître dans un nouveau sous-menu de "Article" dans l'administration de WordPress, à côté des sous-menus "Catégories" et "Mots clés".

La question à se poser est de savoir comment elles ont été conçues, et surtout dans quel but. En référencement naturel, il existe deux cas de figures différents pour les custom taxonomies :

- une taxonomie de "gestion", uniquement utilisée pour classer et ordonner des contenus sans donner accès aux visiteurs à un nouveau type d'archives.
- une taxonomie réelle, à savoir présenter un nouveau formats de liens pour des contenus spécifiques.

### La taxonomie de "gestion"

Voici un exemple sur le site de SeoMix. La page ci-dessous affiche les articles du site Internet selon une taxonomie. La taxonomie en tant que telle n'est pas affichée et celle-ci sert donc surtout à ordonner et classer des contenus : <http://www.seomix.fr/guide-referencement-naturel/>

Si votre taxonomie ne sert qu'à classer des contenus ou à administrer votre site, vous devez donc vous assurer que votre thème ne fait jamais de lien vers vos archives de taxonomies. En règle générale, c'est déjà le cas mais vérifiez quand même vos fichiers de thème WordPress.

Par contre, ce qui peut réellement vous nuire à l'inverse sont les plugins installés. Nous allons donner un simple exemple : si vous avez installé le plugin **WordPress SEO**, celui-ci va par défaut ajouter toutes les taxonomies dans le fichier sitemap. Assurez-vous donc de désactiver les archives de vos taxonomies (ainsi que celles par auteur si vous ne les utilisez pas).

### La taxonomie réelle

Si votre taxonomie doit servir aux internautes, par exemple une taxonomie "films" pour afficher vos contenus *via* une page d'archive dédiée du même nom, vous prenez le même risque qu'avant : avoir un contenu dupliqué entre l'article, la page de taxonomies, les mots clés et les catégories. Il faudra donc utiliser pour chaque page de taxonomie la même méthodologie :

- Tronquer le contenu.
- Ajouter du contenu unique, comme une description de taxonomie avec le code suivant: 

```
<?php $termDescription = term_description( '', get_query_var( 'taxonomy' ) );if($termDescription != '') : ?><?php echo $termDescription; ?><?php endif; ?>
```
- Evitez d'associer le même article à trop de taxonomies (que ce soit vos catégories, vos mots clés ou vos custom taxonomies).

## Conclusion

Vous l'aurez sans doute compris, WordPress possède encore quelques lacunes pour le référencement naturel. Certes, le contenu dupliqué ne va pas vous bannir de Google mais cela peut vous empêcher de grappiller quelques positions de plus sur les mots clés que vous ciblez.

Pour contrer cet effet négatif, il faudra donc agir tant sur le code de votre thème et de vos plugins que sur la gestion interne de vos contenus par les différents rédacteurs de votre site. Et tout devrait bien se passer...

**Daniel Roch**, *Consultant WordPress, Référencement et Webmarketing chez SeoMix* (<http://www.seomix.fr/>)