



Barcamp Nairobi '08
Saturday, June 21 @ Jacaranda Hotel

Wordpress Optimization Tips – Steps to optimize your blog

FINAL

I. Choose the most appropriate title for every single post (IMPORTANT)

By far the most important of all, the title of your articles... You must have major/popular keywords in your title, at least three.

Bad titles could be “*I had fun today*”, “*Just came back from Barcamp*” or “*Barcamp Nairobi*”.

A good/better title would be “*Barcamp Nairobi (Kenya) 2008 – A Conference for African bloggers*”

You can use some Keyword Suggestion Tools (like [Adword Keyword Tool](#) from Google or any free marketing tools such as the one at [DigitalPoint](#)) to get most searched keywords.

II. Permalinks (or nice URLs/URL Rewriting) (IMPORTANT)

By default WordPress uses web URLs which have question marks and lots of numbers in them, however WordPress offers you the ability to create a custom URL structure for your permalinks and archives. This can improve the aesthetics, usability, and forward-compatibility of your links. A [number of tags are available](#), and here are some examples to get you started.

It is important to change the URL displayed by Wordpress so that it contains keywords (e.g. from this [myblog.com/?p=56](#) to this [myblog.com/cat_kw/article_kw](#) or [myblog.com/article_kw](#)). There is really no need to add the date ([myblog.com/2008/05/31/...](#)), the most important being to have keywords in your URL.

NOTE – Do not change this setting if you have already published some articles in your blog unless you know how to deal with XML sitemap, 301 redirect, etc., to remove old URLs from search engines and replace it with new URLs (see sections VII & VIII).

NOTE2 – `mod_rewrite` should be enable in *Apache* for redirect to work with *Wordpress*.

III. Add some Metadata

Make sure that you have added some metadata, the one located in the `<header>` section of your page. The *Meta Description* is sometime used by search engines to display description of your website in their page (usually the description is: *the title of your blog + title of your page*). The *Meta Keywords* is less important that the *Meta Description* because spammers have overused it but it is always good to have a short list of main keywords (usually the keywords should be: *keywords from title of your blog + keywords from title of your page + tags of your article*).

The basic rule is to have at least a different description and article for every single page.

There are plenty of plugins available to assist you, the most popular being [All-In-One SEO Pack](#).

Barcamp Nairobi '08 is an unconference made up of of technical professionals, Internet enthusiasts, bloggers, designers and other clever people in the Nairobi area who wish to share and learn in an open environment.

Barcamp Nairobi '08 is sponsored by: Google Kenya ([google.co.ke](#)), Ushahidi ([ushahidi.com](#)), Strategiclee ([strategiclee.com](#)), BugLabs ([buglabs.net](#)), O'Reilly ([oreilly.com](#)), Yahoo! ([yahoo.com](#)), WordPress ([wordpress.com](#) & [.org](#)), Wananchi, Deep Space Hosting



IV. Write good, structured and tagged content and excerpt

A. Content is king!! (IMPORTANT)

Every webmaster will tell you “*Content is king*”, so make sure to write unique and quality content. Also, do not hesitate to write your personal opinion on the subject and try, if possible, to finish your article with a query or something controversial, in order to get more comments/reactions.

Avoid dropping many lines of text or doing lengthy copy/paste (and if you do, always put a link to the original article)... Illustrate your text if possible and give space between paragraphs. Some recent studies have shown that the shorter the better: article with more than 20 lines of dense text are not properly read.

D. Polish your HTML structure

Do not hesitate to enhance your content by using bold tags (`` is preferred to ``). You could also add some HTML headers if you have titles/subtitle in your article (from `<h1></h1>` to `<h6></h6>`) – do not abuse of headers (check the source code to make sure that headers are used properly) and modify your template/theme if you can to optimize it.

C. Keyword density (IMPORTANT)

Make sure to repeat your main keywords again and again and in different orders. Having a keyword density of about 20 % for main keyword will boost your rank, but again do not abuse of keywords.

A method I am using to increase keyword density without tempering the text of your article is to provide a gallery of images below each post with title and description for every single picture (see my post on [Banksy](#) - in French - for an example). Such gallery using AJAX has another drastic advantage, it reduces bouncing rates and increases time spent on page.

D. Pictures optimization

Always optimize your picture by filing the *alt* and *title* attributes of the `` HTML tag.
``

E. Tag your post

Tags are very popular in the Web 2.0. It enables you to create « *cloud of tags* » on sidebar and to link articles with similar tags. Furthermore, tags are playing a major role to increase your keyword density and to generate your *Meta Keyword* list.

There are some plugins that allow you to implement tagging system and manage tags, one of the most popular begin [Ultimate TagWarrior](#).

F. Make excerpt of your content or use the `<!--more-->` (IMPORTANT)

It is important to either make an excerpt of your content or use the `<!--more-->` HTML tag provided by *Wordpress* to avoid duplicate issues (see below) between home page and the page with your article.

1. Using the excerpt method

By using the excerpt method, you are sure that content in the home page (showing the excerpt if set correctly in the loop file) will be different than the content of the article itself. Another great advantage of the excerpt is that you can really optimize your text by adding keywords, thumbnails and catchy content for the home page and that it can be later use to provide shorten RSS feed.

2. Using the `<!--more-->` method

Using the `<!--more-->` is the easiest method to minimize risk of duplicate content. Your article displayed at the home page will be break where the `<!--more-->` is placed and a link to read further the article will be added.

G. Reduce external links and promote internal links

Every time you are placing an external link in your page, you are loosing a tiny amount of *Page Ranking* (also known as PR) so it is good to reduce the amount of external links placed in your page.

It is not very important if your home page and/or your article itself have a low PR but if you have a relatively high PR, external links should be monitored, especially in your home page.

If you are using the excerpt method, it is advisable to remove the link or place a “*nofollow*” attribute into the link so it is not taking into account by Google Page Ranking algorithm. If you are using the <!--more--> method, you are stuck because you can not modified how the article looks like in your homepage without modifying the article itself. If it is the case, you should try not to place external links at the beginning of your article (before the <!--more-->).

Also, it is important to choose the most convenient external links’ policy:

- Should I use “*nofollow*”?
- Should I open external links into a new window/tab (“*target: _blank*”) or in the same page or give provide both ??

Note that external links are not that bad and they have also some advantages, indeed often it will bring you comments or some incoming link in return (often by the owner the linked website himself).

Internal links are always good and you must have at least two internal links per article.

V Do Update Services as well as Trackback/Pingback

Make sure that *Update Services* is enabled in your *Wordpress* settings so that you blog is easily indexed by major search engines. Indeed, the *Update services* ping search engines – the one listed in your list – to inform them of any new article(s)/URL(s) published in your blog. You can find a list of services to be pinged in the annex of this document. Personally, I don’t do it or only ping one service (*pingomatic*) because it can be very slow and I trust my *sitemap.xml*.

For trackback/pingback to work, the checkbox “*Allow Pings*” should be checked when writing an article (it is enabled by default). This way, every time that you article is mentioned somewhere in a blog, a trackback will be published in your comment section. Note that this will have no impact on your *Page Ranking* since all links from tracbacks/pingbacks as well as comments have a “*nofollow*” attribute but it will definitely brings you new visitors. Personally, all my articles are ping-disabled because showing 100 trackback (like in some blogs) is just horrible.

VI. Provide feeds your blog, promote your blogs and use social networks

A. Provide an RSS feed of your blog

Make sure you are providing feeds of your blog and that your feed are compliant with *W3C XML Validator*. You can optimize your feed by:

- Using a service like *Feedburner*, *Feedcraft* or *Simplefeed* to enhance the compatibility of your RSS and make sure it is. If you use these services, then you should follow their recommendations so that your *Wordpress* feeds are not used and displayed anymore – note that another drastic and even better method is to use *.htaccess* to redirect all your *Wordpress* feeds to your *Feedburner* feed for example.
- Add an image in your feed as well as your favicon to attract visitors
- Add a URL of the article for every single article
- Add *Feedflares* (by *Feedburner*) at the bottom of your feed

Lastly, it’s up to you to decide if you want to show only a part of your article in your feed or to show the totality of the article (see section IV G. about the excerpt method).

B. Promote your blog and make the buzz

Make the buzz by adding some web-social buttons below your articles so that people can click and vote for your article in the hope that you will get enough clicks to be considered as a “buzz” and being published in home page of popular social websites such as *Digg*, *Del.icio.us*, *YahooMyWeb*, *Reddit*, etc... But make sure you are ready for the *Dig Effect* which can overkill your blog and even worst can get you banned by your host if you exceed the allowed bandwidth.

C. Use social networks to build your own network of friends/followers

Create your own network by registering in popular social network such as *FriendFeed*, *Twitter*, *Facebook* which basically provides a feed of all your activities. There are tons of similar sites and tons of applications that help you to populate your last post in social networks. Some think it is a pure waste of time, others are addicted by such websites, it’s up to you to decide but at least you can register for free and try. Lasly, it is usually easier to create your network if you are using a “*brandable*” name.

VII. Generate a sitemap (IMPORTANT)

A sitemap is an XML file with all the URLs of your articles and it enables major search engine to index your blog in ease. Sitemap – created originally by Yahoo – are VERY important and it is a must to have a sitemap for every single website you are running.

They are many plugins that automatically generate/administrate your sitemap in accordance with *Sitemap Protocol*, *XML Sitemap Format* and *W3C XML*. A sitemap usually looks like:

```
<?xml version="1.0" encoding="UTF-8"?>
<urlset xmlns="http://www.sitemaps.org/schemas/sitemap/0.9">
<url>
<loc>http://www.example.com/</loc>
<lastmod>2005-01-01</lastmod>
<changefreq>monthly</changefreq>
<priority>0.8</priority>
</url></urlset>
```

Once your sitemap is generated, register at *Google Webmaster Tools* and add your sitemap (make sure that your URL is the same than your preferred URL)

VIII. Do not overkill your blog and improve loading speed

A. Reduce the number of plugins

Try to reduce as much as possible the number of plugins and widgets used in your blog as some can drastically slow down loading speed of your blog (too many HTTP request, heavy javascript, bad/slow PHP, bad/slow SQL queries). Always choose the most appropriate plugin and make sure there are up-to-date; and take time to read reviews of the plugin from time to time to spot errors/improvements.

B. Get a good host/server

Make sure that your host is good. The webmasters' say is "*You get for what you are paying for*", meaning that if your host is cheap, then you are presumably on a shared account with another hundred websites sharing the same IP than yours. Try to locate your server, find out if it is a shared or dedicated one and then check its respond time to make sure it is not too slow. If you are using free platform like *Blogger* or *Blogspot*, than there nothing you can do.

C. Follow *YSlow* recommendations

YSlow is a *Firefox* extension created by a geek working at Yahoo. Before you install this extension, you must have the popular *Firebug* extension already installed because *YSlow* is a complement to *Firebug*. *YSlow* will check the performance of your website and will output a report on how to improve it. Read the [author's page](#) to know more about *YSlow* and follow their recommendations.

1. Reduce HTTP requests

- Reduce number of javascript calls (compile all *.js* in a single file and compress it);
- Same for CSS calls (compile all *.css* in a single file and compress it);
- Create CSS Sprites for images such as icons or use *Imagemaps* (although it is not compatible with mobile/PDA).

2. Improve your cache control

- One method to improve your cache in *Wordpress* is to use the very popular plugin called [WP-SUPER CACHE](#);
- Another method is to use the Apache module called *mod_headers* and/or *mod_expire* (see below in section VIII) and to set a far future *Expires header* (NOTE: both module should enable in your Apache settings for this to work)
- Check if your server provide some *Etags*

3. GZip your components

- Self-explanatory, use the *Gzip* if available (*mod_gzip* module for Apache 1.3 or *mod_deflate* module for Apache 2.x)

4. Deal with your CSS and JS

- Minify your CSS and JS (remove blank space, reduce the code, etc.)
- Put any CSS at the top, in the header.
- Put some JS at the bottom of the page so that browser can display HTML before downloading scripts
- Do not call the files numerous time if your page

D. Detect bad bots

They are plenty of bad bots/spiders on the web, some are bad, some are good and some are very very bad – they are the one that fetch your website content at incredible speed and disobey at your *robots.txt* file (see section VII).

A good method to detect bad bots is to use the “*Bad bots trap*” technique – that is to hide a link in your home page (e.g. 1x1 transparent GIF/PNG link or hidden anchor) which goes a page located in directory protected/denied by your *robots.txt*. Good bots would detect the link but would not go at the page as instructed by your *robots.txt*; bad bots would simply ignore your *robots.txt* rules and follow the link... You just have to catch the user_agent and other details viewing the protected page and update your list of bad bots to be banned.

To ban bad bots, refer to the section VIII below.

E. Check your error log files and detect slow SQL queries

The error log file is very useful to detect error that occurred on your server and to resolve server-side error that may slow down or even put your site down. Usually the error log file is largely available. If not, you can contact your host and ask them to configure your *PHP.ini* to create a log directory and files.

Implement a [slow query log](#) – the slow query log can be used to find queries that take a long time to execute and therefore that need some optimization. Contact your host to find out if they can set a slow query log (*my.ini* file). Dealing with slow SQL queries can be difficult but slow queries are often the first reason for slow websites, especially websites with high numbers of visitors.

VII. Deal with the duplicate content issue (IMPORTANT)

In order to increase your PR and not being penalized by *Google* for duplicate content (again), the two following steps are VERY important. Duplicate content is very common issue with *Wordpress* and other Content Management System (CMS); indeed many pages can have similar content – like the archive pages (yearly, monthly, daily), tag pages and category pages – and search engines hate this.

Duplicate content can also happen when you are inconsistent with linking URL or do not have a link policy, meaning that sometimes you are using */page/* and */page* and */page/index.htm* (same page but three different URLs). This problem can also arise when you are providing a print or PDF or PDA/mobi version of your pages.

Lastly, duplicate content can also come from scrapers websites (websites that steal your content) and, more surprisingly, even from your aggregator or syndicate partner (websites that fetch your RSS feed to display content).

A. Some easy steps...

- Implement a strong link policy to only use one type of URL (*www* vs *non-www*, *cat/index.html* vs *cat/*)
- Syndicate carefully by providing RSS content slightly different than your own article (shorter, condensed, etc.) and make sure that a link to your original post is included in your RSS so that *Google* can easily track the original article.
- Go to *Google Webmaster Tools* and set your preferred domain feature
- Follow *Matt Cutts*' recommendations by removing any lengthy footer, copyright notice, etc and making an abstract of it with a link to a more detailed page.

B. Make your own robots.txt

A *robots.txt* is a file placed in your root directory in order to instruct all or specific robots not to index and/or follow some directories or files. Therefore, the robots.txt file is the best way to solve duplicate content issues as it instructs search engines to index only your preferred URL and not to show irrelevant URLs.

An example of my robots.txt file for *Wordpress* is shown in the annex. This robots.txt will make sure that ONLY the home page and articles are indexed by search engines. Note that this is where you can also instruct *Google Image* (or any other specific bots) to index your images for example.

C. Change your Metadata Robots

You must also modify your *Metadata Robots* accordingly. To do so, open the file called *header.php* in your theme and look for the `<meta name="robots">` somewhere in your `<header></header>`. Modify it by replacing it with the following script:

```
// DO SOMETHING AGAINST DUPLICATE CONTENT
$name = get_query_var('name');
if( is_single() || (is_page() && ($name!="archives" && $name!="links" ))
|| (is_home() && $paged<="1" )) {
echo '
<meta name="googlebot" content="index,noarchive,follow" />
<meta name="robots" content="all,index,follow" />
<meta name="msnbot" content="all,index,follow" />';
} else {
echo '
<meta name="googlebot" content="noindex,noarchive,follow" />
<meta name="robots" content="noindex,noarchive,follow" />
<meta name="msnbot" content="noindex,noarchive,follow" />';
}
```

This will instruct robots to ONLY index the paginated home pages, the article pages and the independent pages (such as archives, links). Note that search engines will not index but they will still read them to follow links found in these pages and spread the “link sauce”.

Google Webmaster Tool provides a very useful called Robot Tool to check and verify that your *robots.txt* is working properly and that URLs are indeed blocked as planned. Just drop some URLs in the tool and *Google* will tell you if the URL is blocked or not.

B. Make sure Google got it right

To make sure that you are OK, just do the following experiment:

- Type `site:www.myblog.com` on *Google Search* (make sure you have logged off if you are a *Gmail* user)

Normally, *Google* should only return the home page of your blog and one page for every single article you wrote and nothing more.

If you see this: “*In order to show you the most relevant results, we have omitted some entries very similar to the xxx already displayed. If you like, you can [repeat the search with the omitted results included.](#)*” click on the link and check for URL that should not be indexed by *Google*.

If there is something wrong, then use the *Google Webmaster Tools* to remove specific URLs or directories, or modify your *robots.txt*.

E. Some references

- [Duplicate content due to scrapers](#) – Monday, June 09, 2008
- [Deftly dealing with duplicate content](#) - Monday, December 18, 2006
- Ranking As The Original Source For Content You Syndicate – Wed., May 14, 2008
- [Scraped or Stolen Content: What To Do First](#)

VIII. Have a .htaccess file in your server

You will find below a small list of things to be done with your *.htaccess* file (an example of an *.htaccess* file can be found in the annex of this document). Note that dealing with *.htaccess* can be very difficult and a wrong code can easily break your site, therefore it is very important to read documentation before playing with *.htaccess* and, if possible, to test it on your locale machine or test directory. Lastly, never do a stupid copy/paste when dealing with *.htaccess*.

A. Remove any hotlinking protection

- Check you *.htaccess* file and remove any hotlinking protection so that pictures can be displayed in external sites fetching your feed. If you are more advanced, you can only allow hotlinking from specific websites (mostly syndicate websites such as *Feedburner*, *Netvibes*, *iGoogle*, etc.).

B. Ban bad bots using *mod_rewrite*

- Use the Apache *mod_rewrite* module and *RewriteCond* statement to ban bad bots.... Keep your list up-to-date by investigating bad bots found in your daily access log...

C. Improve your cache control

- Set an expire header for every single file type that can be found in your website. Note that Apache *mod_headers* and *mod_expires* should be enabled for this to work. The aim being to set a far future date for file type that are not updated often such as javascript, css, gif/png/jpeg, pdf, etc... forcing server and user's browser to cache these.

D. Deal with canonization issue and wrong URLs

Canonization is when your website can be accessed by multiple URLs or when that multiple URLs are pointing at the same page with same content. If your blog can be viewed using the following URLs (also called canonical URLs) without redirected users – *www.myblog.com*, *myblog.com*, *myblog.com/*, *www.myblog.com/*, *www.myblog.com/index.html*, *www.blog.com/index.html* – then it means your website is not optimized and there is a slight risk of being penalized by search engine as duplicate content, especially if you have spread these URLs on Internet. In order to avoid penalization by search engine:

- The first step, as said above in section VII A., is to go to *Google Webmaster Tools* and set your preferred domain type;
- The second step is to check if Google has indexed any canonical URLs and use the URL Remove Tool available at *Google Webmaster Tools*;
- Implement some redirects in your *.htaccess* so that:
 - ✓ Either *non-www* URLs are redirected to *www* URLs; either *www* URLs are redirected to *non-www* URLs
 - ✓ Deal with wrong URLs such as URLs with multiple contiguous slashes (*myblog.com//cat//*) or wrong URLs (*.htmlm* instead of *.html*) in order to use ONLY ONE consistent URL (like to redirect *<anything>/index.html* to *<anything>/*)

The basic is that all different URLs should be redirect to a SINGLE URL.

This document was written by Thomas Lieven for the Barcamp Nairobi '08.

If you have any problem issue or remark, do not hesitate to contact the author at lievenke@gmail.com

Annexe I

Services to be pinged by Update Services in Wordpress

<http://rpc.pingomatic.com/>
<http://api.feedster.com/ping>
<http://api.moreover.com/ping>
<http://api.my.yahoo.com/rss/ping>
<http://blogsearch.google.com/ping> RPC2
<http://ping.amagle.com/>
<http://ping.bitacoras.com>
<http://ping.blo.gs/>
<http://ping.feedburner.com>
<http://ping.rootblog.com/rpc.php>
<http://ping.syndic8.com/xmlrpc.php>
<http://ping.webloglot.com/rpc.php>
<http://rcs.datashed.net/RPC2/>
<http://rpc.blogbuzzmachine.com/RPC2>
<http://rpc.blogrolling.com/pinger/>
<http://rpc.icerocket.com:10080/>
<http://rpc.newsgator.com/>
<http://rpc.technorati.com/rpc/ping>
<http://rpc.weblogs.com/RPC2>
<http://topicexchange.com/RPC2>
<http://www.blogdigger.com/RPC2>
<http://www.blogooole.com/ping/>
<http://www.blogoon.net/ping/>
<http://www.blogsnow.com/ping>
<http://www.blogstreet.com/xrbin/xmlrpc.cgi>
<http://www.lasermemory.com/lsrc/>
<http://www.newsisfree.com/RPCCloud>
<http://www.popdex.com/addsite.php>
<http://www.snipsnap.org/RPC2>
<http://www.wasalive.com/ping/>
<http://www.weblogues.com/RPC/>
<http://1470.net/api/ping>
<http://www.a2b.cc/setloc/bp.a2b>
<http://api.feedster.com/ping>
<http://api.moreover.com/RPC2>
<http://api.moreover.com/ping>
<http://api.my.yahoo.com/RPC2>
<http://api.my.yahoo.com/rss/ping>
<http://www.bitacoles.net/ping.php>
<http://bitacoras.net/ping>
<http://blogbot.dk/io/xml-rpc.php>
<http://blogdb.jp/xmlrpc>
<http://www.blogdigger.com/RPC2>
<http://blogmatcher.com/u.php>
<http://www.blogooole.com/ping/>
<http://www.blogoon.net/ping/>
<http://www.blogpeople.net/servlet/weblogUpdates>
http://www.blogroots.com/tb_populi.blog?id=1
<http://www.blogshares.com/rpc.php>
<http://www.blogsnow.com/ping>
<http://www.blogstreet.com/xrbin/xmlrpc.cgi>
<http://blog.goo.ne.jp/XMLRPC>
<http://bulkfeeds.net/rpc>
<http://www.catapings.com/ping.php>
<http://coreblog.org/ping/>
<http://www.lasermemory.com/lsrc/>
http://mod-pubsub.org/kn_apps/blogchatt
http://www.mod-pubsub.org/kn_apps/blogchatter/ping.php
<http://www.newsisfree.com/xmlrpcctest.php>
<http://ping.amagle.com/>
<http://ping.bitacoras.com>
<http://ping.blo.gs/>
<http://ping.bloggers.jp/rpc/>
<http://ping.blogmura.jp/rpc/>
<http://ping.cocolog-nifty.com/xmlrpc>
<http://ping.exblog.jp/xmlrpc>
<http://ping.feedburner.com>
<http://ping.myblog.jp>
<http://ping.rootblog.com/rpc.php>
<http://ping.syndic8.com/xmlrpc.php>
<http://ping.webloglot.com/rpc.php>
<http://ping.weblogs.se/>
<http://www.popdex.com/addsite.php>
<http://rcs.datashed.net/RPC2/>
<http://rpc.blogrolling.com/pinger/>
<http://rpc.pingomatic.com/>
<http://rpc.technorati.com/rpc/ping>
<http://rpc.weblogs.com/RPC2>
<http://www.snipsnap.org/RPC2>
<http://trackback.bakeinu.jp/bakeping.php>
<http://topicexchange.com/RPC2>
<http://www.weblogues.com/RPC/>
<http://xping.pubsub.com/ping/>
<http://xmlrpc.blogg.de/>
<http://rpc.twingly.com/>

Example of .htaccess for a Wordpress blog

Note – Exemplified, just replace *myblog.com* by your domain name.

```

<IfModule mod_rewrite.c>
RewriteEngine on
RewriteBase /
</IfModule>

Options All -Indexes
Options +FollowSymLinks

DefaultLanguage en-IS
AddDefaultCharset UTF-8
ServerSignature Off

# BEGIN - Bad bots
<IfModule mod_rewrite.c>
RewriteCond %{HTTP_USER_AGENT} ^(aesop_com_spiderman|alexibot|backweb|bandit|batchftp|bigfoot) [NC,OR]
RewriteCond %{HTTP_USER_AGENT} ^(black.?hole|blackwidow|blowfish|botalot|buddy|builtbottough|bullseye) [NC,OR]
RewriteCond %{HTTP_USER_AGENT} ^(cheesebot|cherrypicker|chinaclaw|collector|copier|copyrightcheck) [NC,OR]
RewriteCond %{HTTP_USER_AGENT} ^(cosmos|crescent|curl|custo|da|diibot|disco|ditto|spider|dragonfly) [NC,OR]
RewriteCond %{HTTP_USER_AGENT} ^(dripleasydl|ebingbong|ecatch|eirgrabber|emailcollector|emailsiphon) [NC,OR]
RewriteCond %{HTTP_USER_AGENT} ^(emailwolf|erocrawler|exabot|eyenetie|filehound|flashget|flunky) [NC,OR]
RewriteCond %{HTTP_USER_AGENT} ^(frontpage|getright|getweb|go.?zilla|go-ahead-got-it|gotit|grabnet) [NC,OR]
RewriteCond %{HTTP_USER_AGENT} ^(grafala|harvest|hloader|hmvie|hhttp|htrack|humanlinks|ilsebot) [NC,OR]
RewriteCond %{HTTP_USER_AGENT} ^(infony|infonybot|infotekies|intelliseek|interget|jira|jennybot|jetcar) [NC,OR]
RewriteCond %{HTTP_USER_AGENT} ^(joc|justview|jyxobot|kenjin|keyword|larbin|leechftp|lexibot|lftp|libweb) [NC,OR]
RewriteCond %{HTTP_USER_AGENT} ^(likse|linkscan|linkwalker|linspiderguy|lwp|magnet|mag-net|markwatch) [NC,OR]
RewriteCond %{HTTP_USER_AGENT} ^(mata.?hari|memo|microsoft.?url|midown.?tool|miixpc|mirror|missigua) [NC,OR]
RewriteCond %{HTTP_USER_AGENT} ^(mister.?pix|moget|mozilla.?newt|nameprotect|navroad|backdoorbot|nearsite) [NC,OR]
RewriteCond %{HTTP_USER_AGENT} ^(net.?vampire|netants|netcraft|netmechanic|netspider|nextgensearchbot) [NC,OR]
RewriteCond %{HTTP_USER_AGENT} ^(attach|nicerspro|nimblecrawler|npbot|octopus|offline.?explorer) [NC,OR]
RewriteCond %{HTTP_USER_AGENT} ^(offline.?navigator|openfind|outfoxbot|pagegrabber|papa|pavuk) [NC,OR]
RewriteCond %{HTTP_USER_AGENT} ^(pcbrower|php.?version.?tracker|pockey|propowerbot|prowebwalker) [NC,OR]
RewriteCond %{HTTP_USER_AGENT} ^(psbot|pump|queryn|recorder|realdownload|reaper|reget|true_robot) [NC,OR]
RewriteCond %{HTTP_USER_AGENT} ^(repomonkey|rma|internetseer|sitesnagger|siphon|slysearch|smartdownload) [NC,OR]
RewriteCond %{HTTP_USER_AGENT} ^(snake|snapbot|snoopy|sogou|spacebison|spankbot|spanner|sqworm|superbot) [NC,OR]
RewriteCond %{HTTP_USER_AGENT} ^(superhttp|surlbot|asterias|suzuran|szukacz|takeout|teleport) [NC,OR]
RewriteCond %{HTTP_USER_AGENT} ^(telesoft|the.?intraformant|thenomad|tightwatbot|titan|urldispatcher) [NC,OR]
RewriteCond %{HTTP_USER_AGENT} ^(turingos|turnitinbot|urly.?warning|vacuum|vci|voideye|whacker) [NC,OR]
RewriteCond %{HTTP_USER_AGENT} ^(wget|widow|wisenuitbot|wwwoffle|xaldon|xenu|zeus|zyborg|anonymouse) [NC,OR]
RewriteCond %{HTTP_USER_AGENT} ^web(alta|zip|email|enhancer|fetch|go.?is|auto|bandit|clip|copier|master|reaper|sauger|site.?quester|whack) [NC,OR]
RewriteCond %{HTTP_USER_AGENT} ^.*(craftbot|download|extract|stripper|sucker|ninja|clshttp|webspider|leacher|collector|grabber|webpictures).*$ [NC]
RewriteRule . - [F,L]
</IfModule>
# END - Bad bots

# BEGIN Canonization
RewriteCond %{HTTP_HOST} !^www\.myblog\.com$ [NC]
RewriteRule ^(.*)$ http://www.myblog.com/$1 [R,L]
# END Canonization

# BEGIN Redirect htm to html
RewriteRule ^(.*)\.htm$ /$1.html [R=301,L]
# BEGIN Redirect "<anything>/index.html" to "<anything>/"
RewriteRule ^(.*)/index.htm$ /$1/ [R=301,L]

# BEGIN WordPress
<IfModule mod_rewrite.c>
RewriteCond %{REQUEST_FILENAME} !-f
RewriteCond %{REQUEST_FILENAME} !-d
RewriteRule . /index.php [L]

```

```
</IfModule>
# END WordPress

<IfModule mod_expires.c>
ExpiresActive On
ExpiresDefault A0
ExpiresByType image/x-icon A26611200
ExpiresByType application/x-javascript A1814400
ExpiresByType text/css A1814400
ExpiresByType image/gif A26611200
ExpiresByType image/png A26611200
ExpiresByType image/jpeg A1814400
ExpiresByType text/plain A300
ExpiresByType application/x-shockwave-flash A1814400
ExpiresByType video/x-flv A1814400
ExpiresByType application/pdf A1814400
ExpiresByType text/html A300
ExpiresByType text/php A0
</IfModule>
```

Robots.txt

// NOTE – This robots.txt is using wildcards which normally not supported by Robots standard, nonetheless most robots support it. Nonetheless, for robots that do not support wildcards *Disallow: /*?** is the same as *Disallow: /* (meaning disallow for the all website).

```
User-agent: *
# disallow all files in these directories
Disallow: /cgi-bin
Disallow: /wp-admin
Disallow: /wp-includes
Disallow: /contact
Disallow: /wp-content/plugins
Disallow: /wp-content/cache
Disallow: /wp-content/themes
Disallow: /trackback
Disallow: /feed
Disallow: /comments
Disallow: */trackback
Disallow: */feed
Disallow: */comments
Disallow: /category/*/*
Disallow: /2006
Disallow: /2007
Disallow: /2008
Disallow: /*?*
Disallow: /*?
Allow: /wp-content/uploads

User-agent: Googlebot
# disallow all files ending with these extensions
(not really necessary but good as example)
Disallow: /*.php$
Disallow: /*.js$

Disallow: /*.inc$
Disallow: /*.css$
Disallow: /*.gz$
Disallow: /*.cgi$
Disallow: /*.wmv$
Disallow: /*.png$
Disallow: /*.gif$
Disallow: /*.jpg$
Disallow: /*.cgi$
Disallow: /*.xhtml$
Disallow: /*.php*
Disallow: */trackback*
Disallow: */feed*
Disallow: /*?*
Allow: /wp-content/uploads

# allow google image bot to search all images
User-agent: Googlebot-Image
Allow: /*

# disallow archiving site
User-agent: ia_archiver
Disallow: /

# disable duggmirror
User-agent: duggmirror
Disallow: /
```