

Matt Cutts: Gadgets, Google, and SEO

Handling noindex meta tags

by MATT CUTTS on AUGUST 30, 2006
in GOOGLE/SEO

Okay, here's a question. I did the search [congoo] recently and didn't get the home page of [Congoo](http://www.congoo.com/)—why not? If you view the source of <http://www.congoo.com/>, it turns out that they have a noindex meta tag:

```
<meta name="robots" content="noindex, nofollow" />
```

Okay, so Congoo apparently doesn't want their root page to show up in search results pages. Fair enough. But just for fun, I did the search on Ask, Yahoo!, and MSN. Ask doesn't show the root page from Congoo, but Yahoo! and MSN do. MSN shows just a url reference:



But if I click on the Cached link, I get the message “Could not find the requested document in the cache.” So it looks like MSN may handle

FOLLOW ME

- [Twitter](#)
- [YouTube](#)
- [Instagram](#)
- [LinkedIn](#)
- [TED](#)

CATEGORIES

30 days (40)

Android (7)

Books/Magazines (29)

Chrome (25)

Food (6)

Fun (48)

Gadgets/Hack (97)

Games (5)

Glossary/Definitions (2)

Google/SEO (554)

Halloween (19)

How to (60)

iPhone (17)

Leftovers (5)

noindex meta tags by showing a url reference but not any snippet.

Now let's look at Yahoo's result:

YAHOO! SEARCH Web | Images | Video | Audio | Directory | Local | News | Shop
congoo

Search Results

Also try: [congoo premium](#), [congoo nashed](#), [congoo publishers](#) More...

- [Install Congoo Netpass Toolbar](#)
[www.congoo.com](#) Free access to over 300 Premium & Subscription websites.

1. [congoo](#)
[www.congoo.com](#) - 15k - Cached - More from this site - Save
2. [Congoo - Free software download](#)
... Search with [Congoo](#) from any web page. ■ Block annoying pop-ups ... Add [www.congoo.com](#) to th
[www.congoo.com/netpass/install](#) - 11k - Cached - More from this site - Save

Huh. They show a 15K page with a Cached link. I clicked on that link and the cached copy had the noindex meta tag on it.

So based on a sample size of one page, it looks like search engines handle the "noindex" meta tag:

- Google doesn't show the page in any way
- Ask doesn't show the page in any way
- MSN shows a url reference and Cached link, but no snippet. Clicking the cached link doesn't return anything.
- Yahoo! shows a url reference and Cached link, but no snippet. Clicking on the cached link returns the cached page.

Something to be aware of. Personally, I'd prefer it if every search engine treated the noindex meta tag by not showing a page in the search results at all.

Jonathan Nelson August 30, 2006 at 11:11 pm
Matt, I completely agree 100%. There needs to be a stronger push for search engine standards. We have the W3C, but where is the accountability?



[Linux/Ubuntu \(28\)](#)

[Movies/Videos \(27\)](#)

[Music \(6\)](#)

[Personal \(143\)](#)

[Productivity \(32\)](#)

[Television \(3\)](#)

[Travel \(30\)](#)

[Uncategorized \(4\)](#)

[Web/Net \(74\)](#)

[Weblog/blog \(109\)](#)

[META](#)

[Log in](#)

[Entries feed](#)

[Comments feed](#)

[WordPress.org](#)

pip August 31, 2006 at 12:17 am



Sorry Matt, there is another important purpose for NOINDEX:

If I set the robots-tag to noindex, it will prevent msn and yahoo from analyzing information ON my page. So if I do a vanity search and tagged my about page noindex i should not find my pages with a search for my name.

since they don't have information about the content of that page its unlikely you will find them in a regular search. so the situation isn't that bad. but i agree it's good to be aware of it and think about using the robots.txt.

Bockereyer August 31, 2006 at 12:53 am



I think, I'm dumb. If I don't want my pages indexed I put a password on them and require registration before access. As far as I know there is no regular bot who does passwords.

Maurice August 31, 2006 at 1:05 am



Hi Matt

Are you really sure Google doesn't display anything at all

I had a site who had blocked their home page using a robots.txt and Google did the same as MSN ie show the bare url confuse me untill i used site maps and nearly had a heart attack when I saw it was reporting the homepage being blocked

I've fixed it now - I can provide the url if you can go back in time some how and see how Google was handling it.

Teodor Filimon August 31, 2006 at 1:11 am



Hmm, whatever purpose we could find for NOINDEX, one thing's for sure: all search engines should treat it the same.

And for the clarity of code, i think NOINDEX should be interpreted as

NO INDEXING (as far as I'm concerned if my URL is stored then it's indexed no matter how much they analyze from my page.)

Matt Cutts August 31, 2006 at 1:23 am



Maurice, we treat robots.txt differently at Google, see here for the rationale:

<http://www.matcutts.com/blog/q-why-doesnt-my-site-show-in-safesearch-or-do-you-hate-metallica/>

I'm more interested in noindex, but certainly password protection is the safest, Bockereyer.

Tony Ruscoe August 31, 2006 at 1:49 am



Personally, I'd prefer it if Google didn't list any URLs in its index that are **DISALLOWed** by the robots.txt file instead of just listing the URL with no title or description as they do now.

BTW, based on a sample size of two websites, it looks like search engines handle the "disallow" robots.txt directive as follows:

- Google shows url references
- Ask doesn't show the pages in any way
- MSN shows url references
- Yahoo! doesn't show the pages in any way

Since there's no equivalent to NOFOLLOW for robots.txt files, it would appear that there's no way to exclude files from Google's index without using the META tags in the page - and that's assuming it's an HTML page...

Is that correct?

JohnMu August 31, 2006 at 1:56 am



Could you use noindex,nofollow on a page instead of a 404? (say your hoster doesn't handle 404's properly, instead returning a 200 with an ad-filled junk page) If you can make the old pages and add the nofollow,noindex, would that help Google remove the page from the index?

Kilroy August 31, 2006 at 2:00 am



Yahoo seems to respect more robots.txt, so maybe we should stick to this methods to forbid crawls. Of course, it would be impossible to prevent only the homepage from being crawled/indexed with this method.

The strange thing about congoo.com is that they have detailed meta keywords and description on their homepage. If so, why prevent crawling and indexing?

Also, the Congoo homepage apparently still has a PageRank displayed (at least on my toolbar). Maybe the noindex tag is a recent addition?

Dave (Original) August 31, 2006 at 2:08 am



Matt, bit off topic, but how about a post on what parts of *popular* forum software (vBulletin etc) users should block via a robots.txt file?

TheSEOGuru August 31, 2006 at 2:57 am



Hello Matts,

It some thing like your blogs code you are using nofollow in your anchor, googlebot doesn't crawl such link but Yahoo and MSN does, You already checked this for my web site.

Well next if you search <http://www.msn.com/robots.txt> , <http://www.yahoo.com/robots.txt> then you will find that they do not use robots.txt file and do not follow the robots syntaxes mentioned in <http://www.robotstxt.org> for search engines

I feel robots only work for google, I have checked this on lot of my web sites.

Cheers 😊

Maurice August 31, 2006 at 3:49 am



Hi Matt

<http://www.mattcutts.com/blog/q-why-doesnt-my-site-show-in-safesearch-or-do-you-hate-metallica/>

I had a look at that but imer still unclear why robots.txt and no index behave diferently.

With my programming hat on (ime defacto CTO here as awell as SEO) having two methods of doing the same thing that behave subtly diferently is not something I would be happy with – it's adding complexity and risk in my opinion.

What's the problem that noindex is solving that robots.txt can't – would be my question which the post about metalica. doesnt seem to answer.

MoreFiles August 31, 2006 at 5:45 am



I've seen a few rel='external nofollow' that Yahoo followed and indexed recently as well

Danny Sullivan August 31, 2006 at 5:46 am



I'm with Maurice, Matt — why would Google want to treat meta noindex and robots.txt differently. They are both intended to do the same thing — keep pages out of an index. The only reason we have two options is simply because some people can't setup robots.txt files for their sites, which might be within the domains of others. However technically they are implemented, it seems like they should be treated the same way.

My gut tells me most webmasters would prefer that all the search engines not list any pages that use either a robots.txt or meta noindex command.

From a user perspective, I think the technique of showing a link to a site if you can learn about it another way is fine, such as being listed in the Open Directory or from links on the public web to those sites.

The Yahoo implementation of meta noindex is odd — why show a cached page. But I can see a hole here. They might not be actually indexing the page but still caching is since the specific noarchive tag isn't also being used:

<http://help.yahoo.com/help/us/ysearch/crawling/crawling-02.html>

Sounds like summit time! Not only would a standard on how meta robots and robots.txt be handy, but it would also be nice to know if blocking a page also inherently blocks caching.

Ken Norton August 31, 2006 at 5:47 am



I find it curious that Congoo has a noindex tag immediately following an overstuffed meta keywords tag. One could hypothesize that they are not exactly sure what they're doing. I found it ironic that an audio voice asks "what are you missing that others have access to?" Um, basic information about SEO? 😊

Gary August 31, 2006 at 6:06 am



Password protection shouldn't be the golden key to not having a page get indexed, that's just plain ridiculous. If webmasters do not want a page indexed, webmasters should have the right keep it from getting indexed, without adding extra security features. It is their website, they pay for it, they own it. By not following the rules, that just seems to me like an invasion of privacy. If you want your phone number unlisted, phone companies might have it in a database somewhere but posting it in the phone book for everyone to see, well that just wouldn't fly.

I think this brings to light the need for standards more than anything else. There should be a committee of at least 1 representative from every significant search engine that looks at what tags should be supported, how they should be interpreted, and determines if new tags should ever be implemented. Whether you use W3C or create an independent group, I think it would go a long way.

soxiam August 31, 2006 at 6:23 am



Matt. I wonder what you think is the best practices regarding noindex tag on the "site down" message page. If for instance, you don't redirect the URL to sitedown.html when you come to <http://www.domain.com>. AND if you included noindex tag on that page (not wanting Google to index and show site down message on serps), would Google remove domain.com from its index?

JohnMu August 31, 2006 at 7:20 am



Soxiam, if your site is down you should return code 503 (Network unavailable or something like that). That lets the bots know it's a temporary problem and to try again later. Returning a page with code 200 ("OK") is just asking for trouble, even if you use robot meta tags on it.

Alan Perkins August 31, 2006 at 8:49 am



>> I'm with Maurice, Matt — why would Google want to treat meta noindex and robots.txt differently. They are both intended to do the same thing — keep pages out of an index. The only reason we have two options is simply because some people can't setup robots.txt files for their sites, which might be within the domains of others. However technically they are implemented, it seems like they should be treated the same way.

robots.txt applies to all robots, not just search engine robots. It's designed to stop robots *reading* pages, not *indexing* pages. Of course, if a page can't be read then its content can't be indexed ... but (in theory at least) its URL can, as can any other data that is deemed relevant to that URL — such as link text, ODP descriptions, etc.

By contrast, the NOINDEX attribute value of the robots meta tag is designed specifically for indexing search engines to obey. Unlike robots.txt, the content at the URL *must* be read in order for the robots meta tag to be seen, and the NOINDEX instruction specifically means that the URL and its content should not be indexed.

An infinitely large (dynamic) site protected by robots.txt need be hit very little by search engine spiders or any type of robot. An infinitely

large site protected by the robots meta tag could be hit very hard indeed by search engine spiders and all other types of robot. This illustrates the key difference between robots.txt and the robots meta tag, IMO.

robots.txt and the robots meta tag each have their places in controlling robotic access to a site, and those places are subtly different. You may need to use both in order to achieve the effect you desire (and sometimes, even using both, you can't achieve the exact effect).

My current concern over robots.txt is that one of the Googlebots (Adwords-Bot) knowingly and deliberately ignores it! See <http://adwords.google.com/support/bin/answer.py?answer=38197>, which reads "Note: In order to avoid increasing CPCs for advertisers who don't intend to restrict AdWords visits to their pages, the system will ignore blanket exclusions (User-agent: *) in robots.txt files." Surely Google should have e-mailed advertisers to ask them to explicitly allow Adwords-Bot to visit, rather than just make the assumption that the advertisers will agree.

It's one thing for Yahoo to ignore NOINDEX, it's another thing for Google to ignore robots.txt – the oldest, most inviolable search engine standard of them all, IMO.

Bombaywala August 31, 2006 at 10:46 am



How about the impact of mod_rewrite? I have index.html indexed, and later I changed all the extensions from .html to .htm

My site does not even have .html file any more – but I used rewrite rule to redirect all the old .html to new .htm respective files.

Should this hurt in indexing?

Joe August 31, 2006 at 11:49 am



I searched for the noindex tag not there so either matt caused them to be say du or what happened to it...

eConsultant August 31, 2006 at 1:35 pm

Google shows del.icio.us homepage for



[http://www.google.com/search?
hl=en&lr=&q=del.icio.us&btnG=Search](http://www.google.com/search?hl=en&lr=&q=del.icio.us&btnG=Search)

even though del.icio.us has noindex/nofollow tag

meta name="robots" content="noarchive,nofollow,noindex"/

and it's robots.txt says:

User-agent: *

Disallow: /

Allow: /rss

(Sorry for the dup : the meta-tag in the first message is not displayed)

S August 31, 2006 at 4:43 pm



I agree, it is odd how msn and yahoo actually index the no index tag. Well, any how, this issue is not effecting anyone because either way no one is finding a non indexed page since they won't search for it.

Dave (Original) August 31, 2006 at 5:08 pm



Hmmm, when I want something out the norm to happen, I would NEVER leave it up to the SE *to decide for me*. If I don't want a page shown in any SE's, I password protect it.

I shouldn't have to lock my front door at night so people can't come in, but I do and I bet most do.

g1smd August 31, 2006 at 6:47 pm



**** Yahoo! shows a url reference and Cached link, but no snippet.**

Even better than that, Yahoo often shows a title in their SERPs for a page that has a noindex meta tag placed on it.

How did they do that?

Well, I see that they simply use the anchor text from one incoming link (one that is on some external site, that is) as the page title; but only if that other site is "trustworthy" and only if that anchor text is not "click here" or some other low-quality generic text.

g1smd August 31, 2006 at 6:58 pm



**** how about a post on what parts of *popular* forum software (vBulletin etc) users should block via a robots.txt file?**

Been there, done that:

<http://www.webmasterworld.com/forum30/33094.htm> and
<http://www.webmasterworld.com/google/3044757.htm> and several other related threads.

TxRex August 31, 2006 at 8:06 pm



It would be nice if the search engines handled this the same but then again having differences can have it's benefits...

Think of it this way...Ford builds an engine with the distributor at the front of the engine which makes it easy to access...GM builds an engine with the distributor at the back of the engine which makes it harder to access...Another company builds an engine and puts the distributor in the middle of the engine...Maintenance on the Ford is easy and needs to be because you have to do it often...Maintenance on the GM is a pain but you live with it because you don't have to do it too often and the engine seems to perform better...The other company is okay to work on but a pain to find the parts because nobody carries them...At any one time depending on your needs, one of the companies above will meet your expectations...Price of gas goes up...the other company looks good because of mileage...Got to haul something...well Ford has their trucks...Need luxury and dependability...GM says we will build vehicles to fit this market...

So with the search engines...a little variation gives us the ability to choose which engine to use...if we don't like what it produces...we don't use it...if a search engine loses the ability to pull people, then revenue goes down...if they don't change, then they risk going the way of the dinosaur...

The standard is simple...

Use robot.txt to exclude directories

Use meta noindex to exclude pages

Use meta noarchive to exclude caching

Use password to protect directories

Use meta nofollow to allow the search engines to make it around the web quicker

The faster the engines get around the web and update the info (including pulling the dead pages out of the index and supplemental index) is more important than how each handles noindex...Besides...he who handles the info the best will be king of the hill during the time his system performs the best...

Really...even in politics we have Republicans, Democrats and Independents...if we made them all the same, then we would be like China or Iran...Do you really want just one company controlling the info and how it is delivered? In Matt's case, you say yes because you are employed...But for the masses, one should say no because competition brings better more evolved products...

There has to be a difference between King of the Hill and God or we are all in trouble...

Watches August 31, 2006 at 10:00 pm

Now that wouldn't be fun if there was only one way to do things now we have different things doing different stuff!!!!!!!!!!!!



Mike August 31, 2006 at 10:02 pm

I am with Alan. robots.txt should be seen as a "Trespassers will be prosecuted" sign. There is nothing to stop people knowing it is there, they just can't go inside. Area 51 is



off limits, but we all know it is there, and speculation is allowed to exist as to why.

The robots noindex meta tag says "Do not keep a copy of this in the index, and do not report it exists". That is the equivalent of a soldier that visits Area 51 and is told to deny it exists. He has permission to visit Area 51, but he is not allowed to divulge that he knows about it.

SEO Tips August 31, 2006 at 10:42 pm



It is surprising to see that major search engines treat metatags and robots.txt differently. Metatags and robots.txt are components of standard HTML, and not of different flavors of a language. Thus, I too feel there needs to be a stronger push for search engine standards, so that there is no ambiguity in the usage of such components.

Klaus Johannes Rusch September 1, 2006 at 12:30 am



What would be nice is a way to completely exclude an area of a site, or a complete domain, from search indexes. Not just the content of the site but the URL as well.

Despite a Disallow: / in robots.txt search engines, including Google, will often display information about the site.

Tony Ruscoe September 1, 2006 at 7:28 am



Dave (Original) said:

If I don't want a page shown in any SE's, I password protect it.

If you don't want *anybody* to view certain pages, then password protecting them makes sense. However, just because you don't want a page to show up in search results doesn't mean you want to keep it behind a locked door for everyone.

There are many reasons for excluding robots from indexing pages, so password protecting them isn't always an option.

jestep September 1, 2006 at 8:04 am



Yahoo shows sites completely indexed even after they are banned. I think they still have the theory going that by having the largest index (or rather largest looking), it somehow makes them more relevant.

Clint Lenard September 1, 2006 at 12:30 pm



Wow, that security code is hard to see before making a post.

I agree with the above that having no Standards among Search Engines makes things a little more interesting. The findings with NOINDEX is odd though and definately something to keep in mind for future reference.

Thanks for the info!

Mortgage Expert September 1, 2006 at 2:18 pm



We are star studed today. We have Danny and Matt posting comments. MSN uses Yahoo engine eons ago. MSN must have reverse engineer before MSN finally use their own search engine. So, MSN behaves like Yahoo at some parts.

Ken Barbalace September 2, 2006 at 2:33 pm



No index should mean no index. If the web page contains "noindex" in the meta instructions then that page should not be stored in SE's database in any shape form or fashion and it should not be listed in any search results regardless of whether or not it is just a URL and title.

There are lots of very legitimate reasons why one would not want to password protect a page but also wouldn't want it in SE indexes.

At the same time the robots.txt instruction "deny" should also be respected for what it means and SEs have no business requesting any page that is explicitly denied via the robots.txt file.

Now some have posts above commented on this as a privacy issue. I see another more legally powerful concern, that being copyrights. The robots.txt file and robots meta instructions (e.g. "noindex" and "noarchive") are industry standard methods for expressing copyright limitations over protected works to search engines. If a page contains the "noindex" instruction and Yahoo ignores that instruction and displays a copy of the page in their cache, they are violating the spirit of the industry standard convention and the copyright owner should be able to hold Yahoo liable for willfully violating of copyrights.

Another Matt September 2, 2006 at 7:26 pm



I was just wondering why congoo wouldn't want there site indexed. Maybe it is just there index and want other pages to show???

Still doesn't make sense to me.

Dave (Original) September 2, 2006 at 10:09 pm



RE: "There are many reasons for excluding robots from indexing pages, so password protecting them isn't always an option."

=====

Oh, it's always on option, that's why it's called on "option" 😊 You only have to have a user add 1 letter, or number, shown clearly on the page, to gain access and keep bots at bay.

We can jump up and down, rant and rave all we like but I doubt that is going to keep the page(s) out of SEs. I would rather be master of my own destiny, than leave it to the inconsistencies of all the different bots out there.

As the saying goes, you want something done properly, do it yourself.

Igor September 3, 2006 at 5:54 am



Hmmm, I've never seen this before, well actually I never tried to check out something like this, but this is a very

interesting thing.

I have no idea how search engines work, but is it possible that MSN Search shows the mother linke . . . I mean the base link. If it shows <http://www.congoo.com/aboutcongoo> then additionally to this it shows the main domain just to tell something the user . . . like "This is the page which belongs to the domain whatever" . I know this sound like nonsense but this could be.

Also this way it's an eyecatcher because the /aboutcongoo page is indent and will make the user to click exactly on this link, beacuse it looks somehow important

The Adam That Doesn't Belong To Matt

September 3, 2006 at 9:25 am



Okay, that's a weird new captcha thing you've got going on, Matt. I can't help but think that it's one that could easily be be bypassed though, since all anyone would have to do is scrape the page, find the code for the captcha input box, and then find the "Please add" and the two numbers. It'd take a good programmer maybe half an hour to pull that off. (Maybe if they were random images?)

I'm not trying to be a jerk or anything...I'm just trying to save you some grief and hassle, that's all. You probably have a hard enough time filtering out blog spam as it is, and I'd hate to see it get worse on you for your sake.

Anyway, my thought on this whole issue...I'm somewhat surprised no one has brought the idea of detecting a user agent up and presenting (or not presenting) content that way. For example, if the Googlebot shows up on a page, it leaves a pretty distinctive trail behind that's easy enough in most server-side languages to detect and process.

On the rare occasions that I don't want to have something indexed and I don't want to password-protect it, I detect the user agent and 301 redirect the bots to something else. That way, they can do whatever they want somewhere else.

Mind you, I tend to do what Dave does, only I take it one step further. I not only password-protect content I don't want indexed,

but I put it into its own directory, which in turn has an entry in my robots.txt file telling bots to leave it alone, and I also use the meta robots tag to do the same thing. Yes, it's total overkill, but if it works then a little healthy paranoia never hurt anyone.

Jorge September 3, 2006 at 10:47 am



When you search for robots.txt, Yahoo shows the sites robots.txt file.

<http://search.yahoo.com/search?p=robots.txt&fr=FP-tab-web-t500&toggle=1&cop=&ei=UTF-8>

IncrediBILL September 3, 2006 at 5:33 pm



Matt,

The phrase "Ask doesn't show the page in any way" is meaningless as Ask doesn't index a lot of stuff. I have a couple of domains referenced all over the place and Ask shows other sites like links to my domains yet they still haven't indexed those domains in 6 months.

Ask wonders why they're at the bottom of the SE heap, big shock.

-Bill

SEOidiot September 4, 2006 at 7:56 am



I think that i prefer the MSN method from a usability standpoint to be honest....

thanks for the maths lessons now as well 😊

Lilly September 4, 2006 at 7:57 am



I was wondering .. if the noindex meta tag helps webmasters prevent duplicate content .. couldn't there be serious issues in the short-term future for those webmasters that heavily use this meta tag as such a solution?

- Lilly

Michael VanDeMar September 4, 2006 at 10:03 am



Matt,

I have a question somewhat related to this. On of the things blackhatters will do on occasion is structure the page so it gets indexed, but no link to the Cached version of the page will show. An example would be something like this:

<http://www.google.com/search?hl=en&q=info%3A8183.10.mil19j.info>

The reason they do this is to hide their cloaked version, the one they are displaying to the bots, from others who would duplicate their methods.

Since afaiik the only reason for doing this would be deception, wouldn't it be a fairly simple process to check if there is going to be a Cache link, and if not don't display the page? I think that the amount of legitimate websites that would be disadvantaged by this would be minimal.

-Michael

Timothy September 5, 2006 at 12:31 am



is it the same result if a site uses robots.txt instead of using the meta tag in the html ??

Ammon September 5, 2006 at 3:05 am



Quote: "From a user perspective, I think the technique of showing a link to a site if you can learn about it another way is fine, such as being listed in the Open Directory or from links on the public web to those sites."

Except of course that no engine is guaranteeing to remove the said link *the moment* it is removed from where it was accidentally mentioned...

For me, a NOINDEX tag says that spiders are free to visit the page, but may not include it in any index.

By contrast, the ROBOTS.TXT tells the spider it is not even allowed to approach the page. "Don't even think of going there, Mr. Spider"

Therefore, the robots.txt is actually the stronger disincentive, surely, if one were to differentiate (although that was NOT the intended purpose of having two options).

The amount of ridiculous obfuscations, jump links, and robot-blocking procedures required to perform the intended purpose of EITHER method of robots exclusion now is almost obscene.

Jonathan September 5, 2006 at 5:14 am

Nice research Matt! Thumbs up!



walkman September 5, 2006 at 6:07 am

Matt,
is it safe to assume that google will not penalize if one of the two identical pages has the no index tag, even if Google pulls the page?

thanks,



OnlyMe September 5, 2006 at 10:12 am

Can I just ask, why would anyone NOT want their home page (main page of the site) to be indexed? Is there a reason or a purpose of not wanting it in the search engines? Matt any suggestions?



gi September 5, 2006 at 10:24 am

How does google handle duplicate data within the same domain? For example, if I have a blog and an entry spans two categories, such as a categories like "work" and "computers" do I get panelized if the data is unique to my blog? This leads to a bigger question. My wife and I run a small online store, and some of



companies we are distributors for give us marketing information to put on our website. I'm starting to rethink the use of that data because other distributors are using either all or some of the same data on their sites. Is my site penalized for using the marketing information or does google realize that we are a legitimate store?

The Adam That Doesn't Belong To Matt

September 6, 2006 at 5:43 am



Can I just ask, why would anyone NOT want their home page (main page of the site) to be indexed? Is there a reason or a purpose of not wanting it in the search engines?

Intranet site.

Private/copyrighted content where the author wants to protect individual property (I've seen travel writers do this, among other people.)

Site built for offline marketing purposes only.

Those are three that come to my mind off the top.

Metal Dome

September 9, 2006 at 7:03 am



Sometimes, the "Baidu" also does not follow the "nofollow" meta tags... Annoying... ..

Google is doing well in this aspect, I like it.

CherryHintonBlue

September 14, 2006 at 2:05 am



Bringing together some of the duplicate content ideas above, as the webmaster of a bunch of sites which have been penalised due to (we think) duplicate content across them (not within them), my solution has been to NOINDEX duplicate pages on all but one site. It was a huge exercise but was it worthwhile? If Google does use TxRex's nice simple summary of what the standard

should be, then I guess we'll not be considered to have duplicate content any more. What about it, Matt?

Hostcento March 6, 2007 at 12:54 am

Really i am surprised when i look at the comments, really sites bother about noindex for their sites?



Accent Pillows March 7, 2007 at 6:50 am

Totally agree with, Matt! Why show on search result and when you click on it you got nothing ?



TerryG March 29, 2007 at 5:31 am

I agree, it is odd how msn and yahoo actually index the no index tag. I have seen quite a few blogs with this code but some or most private website blogs do not. In fact I have found most of my good blog sites, such as this one, through Yahoo search.



Richly Scented Candles June 12, 2007 at 6:47 am

This is a very usefuk feature. Thanks, Matt!



Chris Kameir July 25, 2007 at 1:40 pm

Would it be better to ALLOW the few bots you WANT and DIS-ALLOW any other?



Maximum Persuasion July 26, 2007 at 12:04 am

This no-index thing is an odd ball. Why implement this when there's a stronger form of security found in the .htaccess file?



I used to employ the services of clickbank but got really pissed when the hidden pages redirected from clickbank still ended up in search engine results- especially Yahoo's! And that's despite having noindex and no follow.

sigh.

Wendy December 5, 2007 at 10:28 am



Hi Maximum Persuasion,

What's the stronger form of security .htaccess file that u mentioned?
Is it much more superior than "noindex" & "nofollow"? I'm a newbie
& i'm concerned that my product download page may be displayed
in search engine results.

Can u should about how to do it in .htaccess ?

Many thanks!

Mark February 25, 2008 at 7:46 am



If you indexed a page on our site with a noindex tag, you
would be in violation of our terms of service.

Alexander February 25, 2008 at 8:52 am



Hi Matt,

We are a web design company and I'm the guy who maintains the
company website.

I had a query about Noindex, Nofollow... we have a few websites
that we built but the owners never maintained them and the
domain expired, hence we put those files in a folder and showed
them on our portfolio, till recently we did not put a Rel= Noindex, No
follow tag on the links in our portfolio, but then we saw google
cache these pages and showing them in search results and we
appearing in serps when people we searching for CD duplication
and stuff, which is something we don't do but that client used to.

After the tags we added google was still recaching the pages... we
tried adding a robot.txt, in those folders (we don't use robot.txt in
the root because competitors access these files and find out what
we are trying to hide... trust me its true.. the Industry out here is
really dirty)... Further I even tried adding the tags on the Index files..

but google has cached some inside pages and excel documents which we are struggling to have removed from the index....

Now to my query... If we add No index No follow on the Index page after google has already cached the inside pages and files... Does google only remove the Index page or also the inside pages...

PS. Webmaster tools also does not help much in removing pages from googles index. We had a website which no one maintained for a while and the spammers had a lot of junk on to forum which made this jobsite come up in results for Adult keywords... So when I started maintainance of the site I just renamed the folder in which the forum was installed, which resulted in a 404 error returning for this URL <http://www.ehirers.com/phpBB2> which even showed in the errors on webmaster tools, but when i requested for the removal of this directory from googles index the request was rejected and the only explanation provided was this because that there was a 404(which is very funny) or robot.txt block or Noindex, No follow tag....

Won't be disappointed if i don't receive a reply... but understanding these No index No follow tags and stuff like this only help a Novice like me become an good SEO / Internet Marketer some day...

Thanks for your time people
Alexander Gounder

mark February 28, 2008 at 9:52 am

I agree. "Noindex" shouldn't even show the page in the serp.



Taylan February 28, 2008 at 3:13 pm

Well, its obviosly say "noindex" 😊



Dan October 3, 2008 at 7:53 am

Hi



I just wanted to share my experience with you of these Meta Robots Tags, which caused me a lot of problems until I discovered they

were being added into the web site by WordPress.

The NOINDEX and NOFOLLOW tags were being inserted by the WordPress because of the Privacy settings, I had inadvertently selected or were there by default on my upload.

Anyway if you'd like to know more and my experience about this I've written it here in these articles:

<http://technicalarticles.co.uk/?p=139>

<http://technicalarticles.co.uk/?p=105>

Hope it might help someone else suffering from this problem.

Dan

Alex Visa October 27, 2008 at 12:26 am



Hi Matt,

just by observing the above, we do not have sufficient data to establish whether MSN or Yahoo treat noindex as they should. The 2 search engines may treat the tag wrong, however I have seen noindex nofollow sites which were showing in Google.

You can never know exactly when the 2 tags have been promoted to production, and it may be this happened a very short while ago.

What Google does in such a case, is that it first deletes the snippet of the specific site and ultimately eliminates the listing completely ... but this may take time.

So if you want the test to be 100% accurate, beside the test you are running now, you should prep a page with index and follow, then allow time for indexation and just after that change tags to noindex ...

See what happens then.

Regards,
Alex

Jenny - Fast Weight Loss

December 4, 2008 at 4:33 am



Contrary to Dan's experience above, my encounter with meta robot tag have been positive so far. The proper employment of these tags helps to make sure that those pages I want to keep private really remained private.

I've been using a lot of meta tag as well as the robot.txt file to tell search engine o keep out of certain files/directories. This is especially useful when you have certain digital products/information stored at those private directories.

Seo Company Delhi

December 18, 2008 at 7:54 am



Today (18-Dec-2008) when I visit to congoo.com, I found no such meta tags and this site is raking very good. Perhaps this metatags may have been used for short time or after indexing of pages. After indexing your site it appears for long time in search engines results. There are some others resons also which may be responsible for indexing a site even after nofollow noindex mata tags.

Seo Company Delhi

December 18, 2008 at 7:59 am



Today (18-Dec-2008) when I visit to congoo.com, I found no such meta tags there and this site is raking very good. Perhaps this meta tags may have been used for short time or after indexing of pages. After indexing a site it appears for long time in search engines results. There are some others reasons also which may be responsible for indexing a site even after the use of nofollow, noindex mata tags.

Dragan7

January 30, 2009 at 3:41 am



I am currently developing new website and one question bumped into my head, I was hoping this could be the right place to seek for an answer.

Example:

<http://www.example.com/pageA> links with dofollow link to <http://www.example.com/pageB>

pageB has noindex meta tag. Will page rank flow from pageA to pageB?

Or, if we put noindex on pages that we don't want indexed, should we also put nofollow to all internal links that are leading to those pages?

Wholesale Denim Jeans

May 28, 2009 at 11:21 pm



I put noindex of a directory into the robot.txt. But all the sub-pages of that directory still get indexed by Google. Perhaps, I've put up the robot.txt too late? Google doesn't seem to listen to the noindex command. Anybody knows for sure?

James Blunt

June 14, 2009 at 9:11 pm



I put noindex tag ACCIDENTALLY on large amount of my website, and Google de-indexed it 😞

Now what can I do to get my pages re-indexed as quickly as possible?

Web Design Bangladesh

July 20, 2009 at 11:34 pm



From your post it seems to me that noindex meta tag has a great effect on search engine. So, it is possible to block my site to show on search result. Thanks for your nice information to share.

Seo

September 14, 2009 at 6:11 am



Hi Matt, Nice article but i have some query regarding this (1) Why Google show only very few back links (for a website) as compare to Yahoo even Google don't show follow links most of time and Can you please resolve my this query that how a web page with zero or very few back links have good PR (3-6) OR rank well on SERP as comparisons to other web pages having good back links (no spam) . Please reply to me i am waiting for your reply and thanks for sharing your knowledge with us .

swegill October 14, 2009 at 2:26 pm



Well, we all know that there are the standards and rules.
But for some reason, we don't have the implementation of
those rules or I would say respect for those rules.

I would also vote for WC3 standards should be followed...

cheers!

Swegill

Tim November 25, 2009 at 2:15 am



Hi Matt,

I run a Web site that was recently blocked from Google's index
without explanation. It's a noncommercial site run as a hobby (i.e.
there are no ads, commerce, 'SEO' type activities or motivation for
them), so I was/am quite at a loss to find out why.

After some non-fruitful tests and attempts to find/guess at the
source of the problem (and changing my search pages to Bing.com
in protest ;-), a message appeared in Webmaster Tools claiming to
have found hidden spammy text as justification for the (or new?)
blocking, suggesting the site has been compromised. "The following
is some example hidden text we found at <http://cexx.org/>: order
viagra discount... (list of similar naughty words here)"

A thorough grep of the server found the page it is most likely
triggering on: as part of the anti-malware research we publish on
the site, one file lists all the URLs and search keywords a particular
ad-popping malware product triggers on. The page is intentional
and legitimate (not compromised), however, you can guess the
types of keywords this pest triggers on (viagra, mortgage...), and on
review, I can fully understand how a bot would classify it as spam. In
any case, it's legitimate data, but I'll agree that it has no place being
indexed in any search engine. If I read the above correctly, adding
the NOINDEX meta tag should at least keep this from appearing in
Google and a few other engines.

My question is - although NOINDEX'd pages are not shown in the
index at all, is the page content still read/parsed by Google for other

purposes internally (e.g. the autobanhammer), or does Google essentially abort the HTTP transfer upon encountering the noindex tag? More to the point, is noindex an appropriate way to exclude content from the index that the Algorithm would (either rightly or questionably) use for site classification or assess penalties for?

PS. If you are still maintaining this page, the answer to this question would be helpful incorporated into the discussion above on how Google handles pages with noindex tags. I'm sure I'm not the first/only person to wonder about this!

Thanks
Tim

Dmitry Makarov December 21, 2009 at 11:10 am



Want to publish my email on the webpage, clients complain they cant contact me, but will think twice now before I do that...

Damu April 22, 2010 at 9:33 pm



I am now clear on noindex but still not in nofollow. I have a blog and i don't want search engines to find some pages. Those pages have all been indexed, now i want to add **noindex, nofollow**. Is this nofollow effective on that page only or all the links on the page. I did some search and found few blogs who says it affects all the links on that page as well.

Good explanation on nofollow would be gr8.

Thanks.

Vincent June 18, 2010 at 5:36 am



Hello Matt
if I add the code disallow:pagename in the robots.txt then search engine also not index that page is this right?

Regards
Laxmi Narayan

Carpet London March 31, 2011 at 8:10 am

Can anybody please explain me the meaning of "noindex, nofollow" in meta tags?



Lucien den Arend July 16, 2011 at 2:26 am

Matt,
Where do you say something on the meta tag
'classification'?
Lucien



[← Five Tips for Burning Man](#)

[Scoble visiting the Plex →](#)

[RETURN TO TOP OF PAGE](#)

COPYRIGHT © 2023 · [MATT CUTTS THEME](#) ON [GENESIS FRAMEWORK](#) · [WORDPRESS](#) · [LOG IN](#)