Coding Error Causes Google to Ignore Noindex Directive by @martinibuster

A not commonly known coding bug that’s easily overlooked will cause Google to ignore robots noindex meta tags. This can happen when copy and pasting code from a non-coding environment (like a document editor or a web page). A website was discovered that has a noindex,no follow directive that Google is ignoring, resulting in the web page being indexed.

Screenshot of the web page indexed by Google

The indicated page has a noindex meta tag in the head section of the code. Yet as you can see in this screenshot,the web page

is fully indexed by Google. This may have been caused by an HTML coding error that was introduced by copy and pasting code from a word editing document.

Here is a screenshot of the code:

Screenshot of the code with the error. This error caused Google to ignore a noindex directive. This resulted in a page not intended to be indexed to be fully indexed anyway.

It’s not entirely apparent what is wrong with the code unless you know what to look for.

Curly quotation marks are a difficult to spot coding error that can cause a multitude of problems.

If the administrator of that web page had run the code through the W3C HTML validator, they would have been alerted to a problem:

W3C HTML Validation Service is a free tool that is invaluable for spotting hidden coding errors.

As you can see, the W3C validator instantly caught this problem and flagged it as an error, in red. **Any W3C validator warnings that are flagged in red deserve extra attention and should be reviewed.**

Is this a Google Indexing Bug?

This is not a Google indexing bug. It’s not an issue with how Google indexes websites. This is a coding bug withing the HTML of a web page that is introduced by whoever is cutting and pasting code.

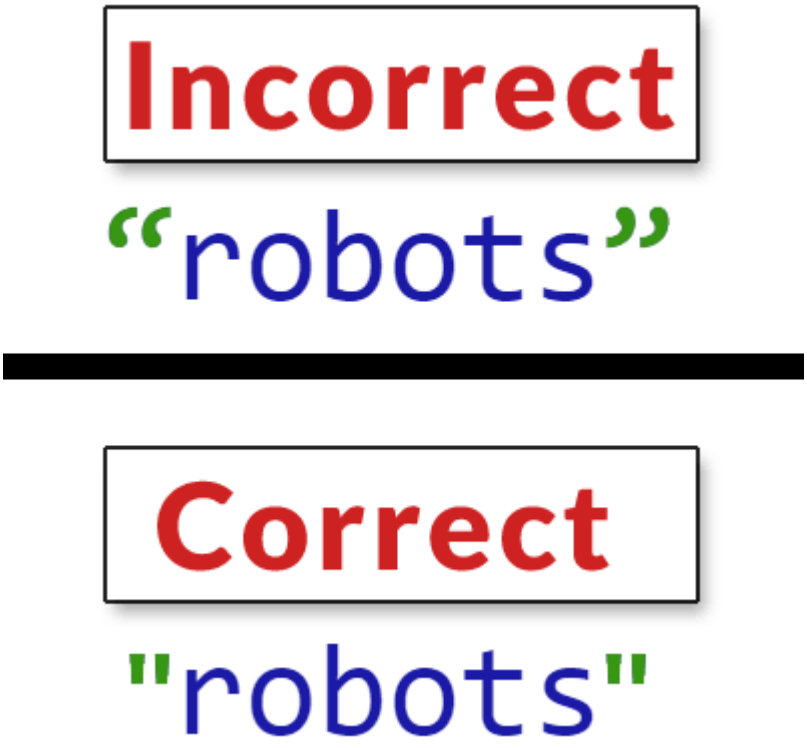
It may be useful if Google added error correction to catch these kinds of errors. But Google is not doing that at this time. So it's on you to practice good coding.

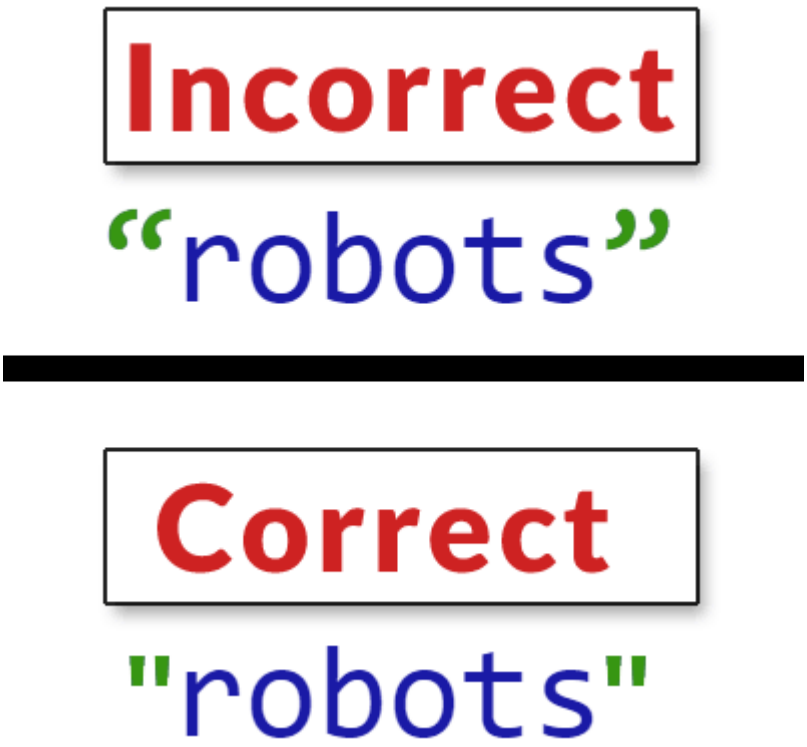
W3C Validator Will Spot this Coding Error

Many people do not run their code through the W3C HTML Validator because HTML doesn't need to validate in order to pass with Google. However in this case, using non-standard characters in the HTML results in an unintended consequence.

Even if you don't care if your code validates as valid HTML, it is still worthwhile to use the W3C Markup Validation Service just in case. (HTML Validator here)

Comparison of incorrect code juxtaposed next to correct code:





Curly quotation marks do not belong in HTML code.

It’s possible that this kind of error, using the wrong kind of quotation marks (curly versus straight) can cause other indexing issues.

Edward Lewis, known as PageOneResults in the old school SEO community, called attention to this in his Facebook feed. This is what Edward said about the issue:

> *“It’s an HTML error that unfortunately Googlebot does not have error correction for. And since it’s in the head, it’s really what I would consider a “Fatal Error” although the validator doesn’t state that.*
>
> *It is fatal for that .edu referenced. Both their nofollow and noindex directives are being ignored by Googlebot.”*

Google Recommends Using Straight

Quotation Marks

The curly versus straight quotation marks bug is actually a fairly common coding issue. Google has a developer style guide that addresses this issue.

> *“Most typefaces support two forms of quotation marks and apostrophes: straight marks and “curly” or “typographic” marks.*
>
> *Some tools, like Google Docs, automatically convert straight quotation marks and apostrophes to the curly versions as you type.* ***However, our guidance is to always use straight quotation marks*** *and straight apostrophes, for these reasons:*
>
> - *It’s easy to get the direction of curly quotation marks (especially apostrophes) wrong. Using straight marks avoids this problem.*
> - ***Code requires straight marks.***
> - *Many writers use coding tools (for example, Subversion) for working directly with markup. Because these tools do not produce curly marks by default, the writer must manually insert the marks, which is error prone.”*

Takeaway:

1. Thousands of duplicate pages or thousands of thin content pages suddenly going live because of a simple mistake can have a devastating impact on search engine rankings.
2. Always remember to run your code through the W3C validator before or just after making it live, just in case. You never know what kind of error it will find.

3. Be careful about copy and pasting your code from a web page or document editing program straight into a coding environment. Document editing apps tend to convert straight quotation marks to curly ones. This can cause fatal coding errors.

Subscribe to SEJ

Get our daily newsletter from SEJ’s Founder Loren Baker about the latest news in the industry!

http://tracking.feedpress.it/link/13962/10985107