

INTERNET ARCHIVE

WayBackMachine

http://microformats.org/wiki/robots-exclusion

Go

OCTJANFEB

263 captures

29 Oct 2005 - 6 Apr 2022

200520062007

About this capture

navigation

- Main Page
- Community portal
- Current events
- Recent changes
- Random page
- Help

search

GoSearch

toolbox

- What links here
- Related changes
- Special pages

Robot Exclusion Profile

Table of contents [hide]

1 Robot Exclusion Profile

1.1 Draft Specification 2005-06-18

1.1.1 Authors

1.1.2 Copyright

1.1.3 Patents

1.2 Abstract

1.3 Introduction

1.4 Format

1.4.1 Profile URI

1.4.2 XMDP Profile

1.5 Examples

1.6 References

1.6.1 Normative

1.6.2 Informative

1.6.3 Thanks

1.7 Issues

1.7.1 Precedence

1.7.2 Specificity

1.7.3 Keywords

1.7.4 Suitability as a microformat

Draft Specification 2005-06-18

Authors

■ Peter Janes

Copyright

This specification is © 2004-2005 by the author. However, the author intends to submit this specification to a standards body with a liberal copyright/licensing policy such as the [GMPG](#). See the [GMPG Principles](#) for more details. Anyone wishing to contribute to this effort MUST read those principles, especially those regarding copyright and licensing, and agree to them before contributing.

Patents

The author neither holds nor intends to apply for any patents on anything required to implement this specification.

Abstract

The Robot Exclusion Profile is a reworking of the Robots META tag (and less-standard extensions) as a [microformat](#).

Introduction

The [Robots META tag](#) is used to provide page-specific direction for web crawlers. While being useful in many cases, its page-specific nature means it cannot be used to restrict crawlers from indexing only certain sections of a document. Several attempts have been made to create more granular solutions through various methods but have perceived shortcomings that limit their use; the Robot Exclusion Profile defines a microformat that can be applied to any element or set of elements in a page.

Like other microformats such as [hCalendar](#), the Robot Exclusion Profile defines a set of class names that may be applied to (X)HTML elements. `class` can be applied to almost every (X)HTML element, which means that authors may be as specific or general as they wish in their application. This differs from the similarly-purposed `rel="nofollow"` attribute, which may only be applied to (and does not refer to the content of) a specific inline link. (It is interesting to note that this behaviour is entirely encompassed by the use of `class="robots-nofollow"` on the same element.) Classes are also additive, so multiple values can be specified at once, e.g. `class="robots-nofollow robots-noindex"`. For robot exclusion in particular, this allows authors to specify multiple rules for an element without adding unnecessary extra markup.

Format

Profile URI

<http://example.org/xmdp/robots-profile#> (obviously preliminary)

The classes defined by the Robot Exclusion Profile should be considered meaningless when the profile URI is not present in the document `<head>`'s `profile` attribute.

XMDP Profile

```
<dl class="profile">
  <dt id="robots-nofollow">robots-nofollow</dt>
  <dd>
    Informs robots that links contained by the element are not to be followed.
  </dd>
  <dt id="robots-follow">robots-follow</dt>
  <dd>
    Informs robots that links contained by the element are to be followed.
  </dd>
  <dt id="robots-noindex">robots-noindex</dt>
  <dd>
    Informs robots that the content of the element is not to be included as part of the page.
  </dd>
  <dt id="robots-index">robots-index</dt>
  <dd>
    Informs robots that the content of the element is to be included as part of the page.
  </dd>
  <dt id="robots-noarchive">robots-noarchive</dt>
  <dd>
    Informs caching robots that the content of the element is not to be included in their cached copy.
  </dd>
```

American Airlines/Hilton, Ex. 1016, p.1
American Airlines/Hilton v. R2 Solutions
IPR2023-00690

```
<dt id="robots-archvive">robots-archvive</dt>
<dd>
  Informs caching robots that the content of the element is to be included in their cached copy.
</dd>
</dl>
```

[\[edit\]](#)

Examples

Removing page content:

```
<head profile="http://example.org/xmdp/robots-profile#">
...
<div class="robots-noindex">There once was a man from Nantucket...</div>
<p>This page is not about <span class="robots-noindex">pornography</span>.</p>
```

Showing `nofollow` in conjunction with `votelinks`, and applying it in parallel with `relnofollow`:

```
<head profile="http://example.org/xmdp/robots-profile#">
...
<p class="robots-nofollow">This is <a href="http://example.com/bogus">a bogus link</a>
and so is <a href="http://example.net/bogus">this</a>.</p>

<p>I don't like <a rel="nofollow" rev="vote-against" class="robots-nofollow"
href="http://example.com/disagree">this page</a>
but I do like <a rev="vote-for" href="http://example.com/agree">this one</a>.</p>
```

Preventing images from being stored by search engines, forcing them to be retrieved from the originating website:

```
<head profile="http://example.org/xmdp/robots-profile#">
...
<p></p>
```

A [more complex example](#)  is available which also shows how the robots metadata may be [visualized](#) .

[\[edit\]](#)

References

Normative

[\[edit\]](#)

- [XMDP](#) 
- [The Robots META Tag](#) 

Informative

[\[edit\]](#)

- [A Standard for Robot Exclusion](#) 
- [Googlebot Frequently Asked Questions](#) 
- [The ROBOTS META Tag](#) 
- [RelNoFollow Draft Specification](#)
- This page was contributed from the [technorati developers' wiki](#) .

Thanks

[\[edit\]](#)

- [Tantek Çelik](#) 
- [Lachlan Hunt](#) 
- [Joe D'Andrea](#) 

Issues

[\[edit\]](#)

These are open issues that have been raised in various forums. The "efficacy" and "collateral damage" issues from `rel="nofollow"` also apply.

Precedence

[\[edit\]](#)

- Should earlier values take precedence or later? Does `class="robots-nofollow robots-follow"` means the same as `class="robots-nofollow"` Or `class="robots-follow"`?
- `meta` tag suggests not using conflicting or repeating directives and so does not specify precedence
- Interaction with [relnofollow](#): what does `class="robots-follow" rel="nofollow"` mean? Currently [relnofollow](#) has no profile URI defined, so the Robot Exclusion Profile takes precedence. In the future, per XMDP's [Using Multiple Profiles](#) , <q>the URIs in the 'profile' attribute are to be treated most significant (first) to least significant (last).</q>

Specificity

[\[edit\]](#)

- Does not allow control of specific UAs à la [A Standard for Robot Exclusion](#) 

If it is actually necessary to control specific UAs here is an possible soluiton, Example:

```
<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Strict//EN" "http://www.w3.org/TR/xhtml1/DTD/xhtml1-strict.dtd">
<html>
<head>
<link rel="schema.RobotExclusion" href="http://example.org/..." />
<meta name="RobotExclusion.RobotName1" content="Foo Bot" />
<meta name="RobotExclusion.RobotName2" content="Bar Bot" />
<meta name="RobotExclusion.RobotName3" content="Evil Bot" />
</head>
<body>
<h1>Page</h1>
<p class="robots-noindex">This paragraph shouldn't be indexed by any bot.</p>
<p class="robot3-noindex">This paragraph should be indexed by every bot except "Evil Bot".</p>
<p class="robots-noindex robot1-index">This paragraph should be indexed by any bot except "Foo Bot".</p>
</div>
</body>
</html>
```

Of course it is a waste of bandwidth if there are "RobotExclusion.RobotName" meta tags on every page of a website. Thus this metatags should be stored on one page - perhaps the main page - so they can be maintained easily.

```
<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Strict//EN" "http://www.w3.org/TR/xhtml1/DTD/xhtml1-strict.dtd">
<html>
<head>
<link rel="schema.RobotExclusion" href="http://example.org/..." />
<link rel="RobotExclusion.Names" href="http://mypage.com/" />
</head>
<body>
<h1>Page</h1>
<p class="robots-noindex">This paragraph shouldn't be indexed by any bot.</p>
<p class="robot3-noindex">This paragraph should be indexed by every bot except "Evil Bot".</p>
<p class="robots-noindex robot1-index">This paragraph should be indexed by any bot except "Foo Bot".</p>
</div>
```

```
</div>
</body>
</html>
```

Keywords

[\[edit\]](#)

- The keywords `all` and `none` are defined by the Robots META Tag as convenience shortcuts to enable or disable the combination of `nofollow` and `noindex`, but predate Google's `noarchive` and should not be considered to include it. As a result, for purposes of clarity and simplicity (the [XMDP Minimalism principle](#) 🗿), they are not included in this version of the Robot Exclusion Profile.

Suitability as a microformat

[\[edit\]](#)

- Isn't the Robot Exclusion Profile designed for machines first and humans second instead of vice versa? Yes, just as much as [relnofollow](#), the deployed microformat that it's designed to replace.

This page was last modified 16:19, 5 Nov 2005.

This page has been accessed 2275 times.

[About Microformats](#)

[Disclaimers](#)

