Useful HTML Meta Tags

These are the HTML **Meta Tags** that I find useful or interesting. I am not intending to document all possible Meta Tags here. Check the references for more detail and other Meta Tags.

Meta Tags References

HTTP 1.1 RFC 2068
Vancouver Webpages on Meta Tags
Workshop Report on Spidering
Caching Tutorial for Web Authors and Webmasters
Remove a Site from Google

XENCRAFT

Web and Globalization Services

I18nGuy Home Page

Useful META TAGS Table of Contents

| Author | Cache-Control | Content-Language | Content-Type | Copyright | Description | Expires | Googlebot | Keywords |

| Pragma No-Cache | Refresh | Robots |

Note the keywords "HTTP-EQUIV", "Name" and "Content" are case-insensitive. Their values are also case-insensitive.

TAG NAME	EXAMPLE(S)	DESCRIPTION
AUTHOR	<META NAME="AUTHOR" CONTENT="Tex Texin">	The author's name.
CACHE-CONTROL	<META HTTP-EQUIV="CACHE-CONTROL" CONTENT="NO-CACHE">	HTTP 1.1. Allowed values = PUBLIC | PRIVATE | NO-CACHE | NO-STORE. **Public** - may be cached in public shared caches **Private** - may only be cached in private cache **no-Cache** - may not be cached **no-Store** - may be cached but not archived The directive CACHE-CONTROL:NO-CACHE indicates cached information should not be used and instead requests should be forwarded to the origin server. This directive has the same semantics as the **PRAGMA:NO-CACHE**. Clients SHOULD include both PRAGMA:NO-CACHE and CACHE-CONTROL:NO-CACHE when a no-cache

		request is sent to a server not known to be HTTP/1.1 compliant. Also see EXPIRES. **Note:** It may be better to specify cache commands in HTTP than in META statements, where they can influence more than the browser, but proxies and other intermediaries that may cache information.
CONTENT-LANGUAGE	`<META HTTP-EQUIV="CONTENT-LANGUAGE" CONTENT="en-US,fr">`	Declares the primary natural language(s) of the document. May be used by search engines to categorize by language.
CONTENT-TYPE	`<META HTTP-EQUIV="CONTENT-TYPE" CONTENT="text/html; charset=UTF-8">`	The HTTP content type may be extended to give the character set. It is recommended to always use this tag and to specify the charset.
COPYRIGHT	`<META NAME="COPYRIGHT" CONTENT="© 2004 Tex Texin">`	A copyright statement.
DESCRIPTION	`<META NAME="DESCRIPTION" CONTENT="...summary of web page...">`	The text can be used when printing a summary of the document. The text should not contain any formatting information. Used by some search engines to describe your document. Particularly important if your document has very little text, is a frameset, or has extensive scripts at the top.
EXPIRES	`<META HTTP-EQUIV="EXPIRES" CONTENT="Mon, 22 Jul 2002 11:12:01 GMT">`	The date and time after which the document should be considered expired. An illegal EXPIRES date, e.g. "0", is interpreted as "now". Setting EXPIRES to 0 may thus be used to force a modification check at each visit. Web robots may delete expired documents from a search engine, or schedule a revisit. HTTP 1.1 (RFC 2068) specifies that all HTTP date/time stamps MUST be generated in Greenwich Mean Time (GMT) and in RFC 1123 format. RFC 1123 format = **wkday "," SP date SP time SP "GMT"**

		wkday = (Mon, Tue, Wed, Thu, Fri, Sat, Sun) date = 2DIGIT SP month SP 4DIGIT ; day month year (e.g., 02 Jun 1982) time = 2DIGIT ":" 2DIGIT ":" 2DIGIT ; 00:00:00 - 23:59:59 month = (Jan, Feb, Mar, Apr, May, Jun, Jul, Aug, Sep, Oct, Nov, Dec)
KEYWORDS	`<META NAME="KEYWORDS" CONTENT="sex, drugs, rock & roll">`	The keywords are used by some search engines to index your document in addition to words from the title and document body. Typically used for synonyms and alternates of title words. Consider adding frequent misspellings. e.g. heirarchy, hierarchy.
PRAGMA NO-CACHE	`<META HTTP-EQUIV="PRAGMA" CONTENT="NO-CACHE">`	This directive indicates cached information should not be used and instead requests should be forwarded to the origin server. This directive has the same semantics as the CACHE-CONTROL:NO-CACHE directive and is provided for backwards compatibility with HTTP/1.0. Clients SHOULD include both PRAGMA:NO-CACHE and CACHE-CONTROL:NO-CACHE when a no-cache request is sent to a server not known to be HTTP/1.1 compliant. HTTP/1.1 clients SHOULD NOT send the PRAGMA request-header. HTTP/1.1 caches SHOULD treat "PRAGMA:NO-CACHE" as if the client had sent "CACHE-CONTROL:NO-CACHE". Also see EXPIRES.
REFRESH	`<META HTTP-EQUIV="REFRESH" CONTENT="15;URL=http://www.I18nGuy.com/index.html">`	Specifies a delay in seconds before the browser automatically reloads the document. Optionally, specifies an alternative URL to load, making this command useful for **redirecting** browsers to other pages.
		CONTENT="ALL | NONE | NOINDEX | INDEX| NOFOLLOW | FOLLOW | **NOARCHIVE**"

| ROBOTS | `<META NAME="ROBOTS" CONTENT="ALL">`

`<META NAME="ROBOTS" CONTENT="INDEX,NOFOLLOW">`

`<META NAME="ROBOTS" CONTENT="NOINDEX,FOLLOW">`

`<META NAME="ROBOTS" CONTENT="NONE">` | default = empty = "ALL"
"NONE" = "NOINDEX, NOFOLLOW"

The CONTENT field is a comma separated list:
INDEX: search engine robots should include this page.
FOLLOW: robots should follow links from this page to other pages.
NOINDEX: links can be explored, although the page is not indexed.
NOFOLLOW: the page can be indexed, but no links are explored.
NONE: robots can ignore the page.
NOARCHIVE: Google uses this to prevent archiving of the page. See http://www.google.com/bot.html |
| GOOGLEBOT | `<META NAME="GOOGLEBOT" CONTENT="NOARCHIVE">` | In addition to the ROBOTS META Command above, Google supports a GOOGLEBOT command. With it, you can tell Google that you do not want the page archived, but allow other search engines to do so. If you specify this command, Google will not save the page and the page will be unavailable via its cache. See **Google's FAQ**. |