



MetaTXT: Solving Mobile Search through Simple Technology

Author: Niall Cusack,
CTO
visibility mobile
Version: 1.0
Date: 9th September 2008

Foreword

The World Wide Web is built upon a set of technologies chosen for their simplicity. This simplicity was a key factor in growing the Web from one site and a handful of pages in 1992 to over 100 million sites and tens of billions of pages in 2008.

The assumption since the Web's inception is that each page should be viewable in all browsers. With the inclusion of browsers on mobile handsets, this expectation is no longer reasonable. Although certain mobile browsers, such as the iPhone's Safari browser, the WebKit-based browser on the Nokia S60 phones and the latest Opera Mobile browsers, do allow Web pages to be viewed as they were intended, these innovative browsers do not change the fact that mobile handsets will always have severe screen size limitations (the largest just 8% of the size of a typical PC screen). Also, the vast majority of handsets have just 15 keys (compared to 100+ key PC keyboards), and at best a tiny trackball for navigation.

Because of these differences, all of the most popular Web sites, along with tens of thousands of other new and existing web sites, have decided it best to create mobile-optimized versions of their sites. These modified sites provide an experience better tuned toward the small handset screens, limited input capabilities and use cases for people on the go. However, in creating these sites, a gap has appeared in the technologies underlying the Web. Specifically, given only the address of a Web site on a PC, how do people using mobile devices find the mobile-optimized content?

Taking a step back from that question, one realizes this is not a new problem, but instead part of a broader issue that appeared with the invention of RSS feeds, podcasts, and other forms of content which are not traditional web "pages." The more general question is, given only the address of a web site, how does a user find all the variations of content related to that site, beyond the PC-optimized pages?

Fortunately, the underlying web technologies were valued for their simplicity, which allows for the implementation of simple solutions to unforeseen problems such as this. For example, Web crawlers were not part of the original Web technologies, but the robots.txt file was quickly implemented to control these crawlers. After many years, this proved to be only a partial solution, and sitemaps were created to further optimize Web crawling. In another example, as people's lists of Web bookmarks grew from a handful into hundreds, the "favicon" was invented to help visually differentiate bookmarks.

It is from these examples that the meta.txt file has been modeled. The meta.txt file is a simple text-file solution to the problem of finding a site's content and variations. Taking less than five minutes to generate for most webmasters, meta.txt can save countless hours for both site creators and visitors, while creating a better experience for billions of Web surfers.

Michael Luni Libes
Chief Architect and Co Founder
Medio

Meta.txt – A Standard for Improving Mobile Web Browsing and Discovery of Media

Executive Summary

Recently there has been a surge in articles detailing just how brands and mobile websites can become more visible on the Mobile Internet. From rules, to 10 point plans, to pages of best practices – the PC web is awash with advice. Many attempts have been made to simplify mobile website discoverability, however so far the responsibility has been on the webmaster to create a site that fits an ever increasing list of criteria – remember the user is in a rush; the screen on a mobile is a lot smaller than that of a PC; join a mobile site directory. No one has put much focus on a possible technology solution. Until now.

Discoverability has proved difficult because most search engines point the user to the PC URL first, and from there the user is pointed to the mobile version, if it exists. This is slow for the user and does not reflect the importance of having an up to date and user friendly mobile site – the PC site always take precedence. With an expected surge in the number of mobile web users in the coming years, it becomes vitally important to improve the end user experience.

This white paper introduces a new standard, meta.txt, which not only directs users to the site most suitable to their platform but will also help users to find results more relevant to their geographical location. All of this is done regardless of domain name extension of a website. Furthermore, meta.txt allows for improved discoverability of media such as rss, podcasts and videocasts. These features combined will allow users to retrieve information/media most relevant to them or their location, will improve load times significantly and allow end users to view their results in the best available form.

What is MetaTXT?

Meta.txt defines multiple User-Agent entry points for a suite of web sites, be they the more traditional pc sites, mobile sites or the media that they contain. Acting as a peer to the robots.txt file, meta.txt is a HTTP accessible file which resides on the local URL “/meta.txt”. Working in a converse manner to robots.txt, meta.txt files will indicate which url a User-Agent should follow to bring end users to web site information most relevant to them. Additionally, meta.txt will allow end users to have this information presented to them in the format most appropriate to the platform they are using to access the web, moving us closer to the utopian vision of “One Web” (<http://www.w3.org/TR/mobile-bp/#OneWeb>).

Format of MetaTXT

Following the same colon delimited format of robots.txt, meta.txt becomes an easily parsed, human readable file. The meta.txt file consists of one or more fields terminated by CR,CR/NL, or NL. Each field in the file follows the form <fieldname>:<value>. Fields not adhering to this syntax or fields with unrecognisable fieldnames are ignored.

The major advantage of this format is that meta.txt can be extended to include as of yet implemented forms of browsing. Details of the currently supported fields are listed below.

Meta Information

name:

Currently the name of a site is the same as the title of the index page of a site defined within its <title></title> HTML tags. The value of the name field thus becomes the text defined within the <title> tags.

eg: name:example.com

description:

The description field holds a short summary of a site which is otherwise generated by a search engine. Currently the description of a site is defined using the description meta tag.

eg: description:example.com is a widely used example website

keywords:

Working in the same way as the keywords meta tag the keywords field lists words and phrases that best describe what the majority or most important part of a website's content is about.

eg: keywords:example, demo, demonstration

PC Entry Page

pc:

The pc field holds the home page for the PC-optimised version of the web site.

eg: pc:<http://www.example.com>

Mobile Entry Pages

The mobile entry pages fields define either a single User-Agent entry point or multiple entry points for varying versions of mobile sites. If a meta.txt file does not contain a pc: field then it should contain at least one of the following of the below entry points.

Single Entry Point

mobile:

mobile: stores the home page of a mobile site that is intended for use across all mobile web platforms.

eg: mobile:<http://m.example.com/>

Multiple Entry Points

Each of the entry points defined below stores the home pages for differing versions mobile sites

mobile-smartphone:

eg: mobile-smartphone:<http://example.com/pda>

mobile-wap2:

eg: mobile-wap2:<http://example.com/xhtml>

mobile-wap1:

eg: mobile-wap1:<http://example.com/wml>

Sitemaps

Each of the above home pages/entry points can also include a URL to specify the location of the associated sitemap (<http://sitemaps.org>). This is done by suffixing the element with `-sitemap`. The sitemap fields should contain the absolute urls for a site's sitemap

pc-sitemap:

eg: `pc-sitemap:`<http://example.com/sitemap.xml>

mobile-sitemap:

eg: `mobile-sitemap:`<http://example.com/sitemap.xml>

mobile-smartphone-sitemap:

eg: `mobile-smartphone-sitemap:`<http://example.com/pda-sitemap.xml>

mobile-wap2-sitemap:

eg: `mobile-wap2-sitemap:`<http://example.com/mobile-sitemap.xml>

mobile-wap1-sitemap:

eg: `mobile-wap1-sitemap:`<http://example.com/mobile-wap1-sitemap.xml>

RSS

The following fields contain the urls for various forms of rss feeds that may or may not appear on a site. The fields are broken into text based feeds (rss), audio feeds/podcasts (podcast) and vide feeds/videocasts (video). If a site has multiple feeds the fields may be duplicated.

rss:

eg: `rss:`<http://example.com/rss.xml>

podcast:

eg: `podcast:`<http://example.com/podcast.xml>

video:

eg: `video:`<http://example.com/video.xml>

Location

Location fields enable geotagging of a site and its content. The fields take note of a site's latitude, longitude and country code. This is of great benefit as it both shows where a site/media is located and allows for return of results most relevant to a user's location.

latitude:

eg: latitude:52.25244946111211

longitude:

eg: longitude:-7.18806266784668

region:

eg: region:IE

Example

The below is an example of a meta.txt file for a site called example.com. The file contains details on meta information, pc site url, a single mobile url, three text based rss feeds, one audio/podcast feed, one video feed and tells us that the site is based in Myanmar at the given longitude and latitude.

```
name:example.com
description: example.com is a widely used example website
keywords: example, demo, demonstration
pc: http://www.example.com
mobile:http://m.example.com
rss:http://rss.example.com/rss/topstoriesoftheday.xml
rss:http://rss.example.com/rss/toppoliticalstory.xml
rss:http://rss.example.com/rss/topsportstory.xml
podcast:http://rss.example.com/podcasting/news.xml
video:http://rss.example.com/rss/tutorial.xml
longitude:12.3456789
latitude:98.7654321
region:MM
```

Conclusion

This new standard will improve the usability of the mobile web and now it is up to you the mobile web masters and mobile search engines to adopt this approach and make the Mobile Internet better. For more information about how to support metaTXT or to join the working group shaping the standard please email us at sinead@visibilitymobile.com