STATE OF THE ART HTML CODING
MEETING SEARCH ENGINE OPTIMIZATION STANDARDS

Gabriel Eugen Garais[1]

ABSTRACT

Today's standards of web codding have come to a point in which web semantics must be of great interest in developing webpages. Relevance of content through web semantics is a must in implementing Search Engine Optimization rules. Content that is included in a web page must be part of a category that treats the subject from many point of views and has also extended versions so that the web page gets a high rank of authority for the written and presented subject on the Internet.

Keywords: SEO, HTML, Tags, shared vocabularies, RSS

1. INTRODUCTION

One of the tools that stand at the basis of the subject trends analysis is the Google Keyword Planner. Using this tool you'll be able to see the number of monthly searches on the keywords that you have in mind to include in the web page. This tool will also provide you with ideas on using similar words that have increased search frequency over the Internet on specific periods of time. You can also anticipate the periodicity on querying based on the periods of time and frequency of use. After choosing the right words for expressing the desired subject we must outline some of the keywords and elements in the web page with specific HTML tags so that the machines formed by web servers and search engine crawlers will understand the meaning and importance of the written subject.

2. SHARED VOCABULARIES

Servers do not communicate as humans do so for this connection to be possible there must be some implementations to be made. There is too much information that travels through the Internet and machines must understand and distinguish between different types and meanings of contents. To accomplish this, the Internet now uses semantic web standards. The semantic web relies on standard vocabularies which the most important are Schema.org, RDF, OGP, Dublin Core so on and so forth. This vocabularies standards use specific html tags that must be applied to each content in a Web Page that represents valuable Information for the presented subject. Using metadata tags has its advantages for SEO because search engines interpret the extra code as an effort to make web page content as relevant as possible.

The Dublin Core

According to [1], [2] and [3] The Dublin Core Metadata Initiative (DCMI) is an open source movement that started in Dublin, Ohio, to standardize data about digital objects. It uses a set of metadata elements that complement HTML metadata. The Dublin Core (DC) metadata prepares the web page for the constantly evolving semantic web. It is very simple

[1] Lecturer PhD, Romanian-American University, Bucharest, garais.gabriel.eugen@profesor.rau.ro

and the implementation cost is low. This initiative works for XML, XHTML and HTML pages. Whether or not it helps search rankings has always been under debate but any effort made toward cleaner and more easily understandable code is favored by search engines. Considering how important relevance is in today's world of on-page SEO, the keywords represented in the DC code must be consistently present all over the web page, i.e. in the title, description, heading and so on. Although Google has clearly said that it no longer uses meta keywords information to identify the target keywords of a specific page, some experts still feel that Meta keywords are used by search engines. So, to be on the safe side, instead of using the meta keyword in your HTML code you could use the meta keyword (subject) element in the DC metadata. Advantages of using Dublin Core metadata from a usability point of view: It is easy to implement; It does not cause code bloat; It is part of standard coding, hence it will not become outdated; It helps some internal search engines to improve search and usability within the site.

The following are 22 elements that feature in Dublin Core Metadata:

- **Title:** The page title given to the document.
- **Subject:** The keywords or key phrases used that are relevant to the information found in the title, description and page content.
- **Description:** A text (or graphical) account of the content of the page.
- **Type:** Declaring the type of content found on the page; for example, image, sound or text.
- **Source:** Display the original source of the content, for example, the page number of a printable book.
- **Relation:** Relation can include what the page is a part of, version of, format of, referenced by or based on. For instance, if the Title is *The movie My Fair Lady*, the Relation would be (based on) *Shaw's play Pygmalion.*
- **Coverage:** This element is used to give spatial, geographical or time-based information. Examples of coverage are *Houston, TX, 18th Century* or *2009-2012.*
- **Creator:** This includes the name of the person, organization or service responsible for creating the web page content.
- **Publisher:** This includes the name of the person, organization or service that has made the web page content available.
- **Contributor:** This would include the name of a person, organization or service that has contributed to the content on the web page.
- **Rights:** Intellectual Property Rights (IPR), Copyright and any other resources or services providing rights information must be displayed here. It could contain written information or a URL pointing to a page with rights information.
- **Date:** This could include a date associated with an event in the life of the content on the page. It must be in YYYY-MM-DD format. You could also provide YYYY-MM or just YYYY.
- **Format:** This must include the media type, image type or dimensions of the content of the page. Dimensions may include size or duration. For example, *image/gif, 50 x 512 pixels.*
- **Identifier:** This references the page content by means of a string or number assigned to the content by a formal identification system, such as the Uniform

Resource Identifier (URI)/Uniform Resource Locator (URL), Digital Object Identifier (DOI) and International Standard Book Number (ISBN).

- **Language:** This states the language of the content. You can use a language code or a text string, for example, *en*, *fr* or *Primarily English, with some abstracts in French*.
- **Audience:** Define your audience in this element; for example, elementary school students, deaf adults etc.
- **Provenance:** If there has been a change in the ownership of content since its creation, use this element to describe that. For example, *This copy was previously owned by James Elliot.*
- **RightsHolder:** Use the name or the URI of the rights holder of the content.
- **InstructionalMethod:** This element includes the ways in which the information on the web page is to be used. For example, *Experimental learning, Observation, Large Group Instructions*.
- **AccrualMethod:** This describes the method by which items are added to a collection. For example, *Deposit, Purchase.*
- **AccrualPeriodicity:** This describes the frequency with which items are added to a collection. For example, *Annually* or *Irregularly.*
- **AccrualPolicy:** This describes any policy governing the addition of items to a collection. For example, *Active* or *Closed.*

Schema.org

According to [4] a shared markup vocabulary makes it easier for webmasters to decide on a markup schema and get the maximum benefit for their efforts. Search engines have come together to provide a shared collection of schemas that webmasters can use. Schema.org was launched on June 2, 2011, providing a structured data markup supported by major search engines. According to [5] since the lunch, schema.org markup has been implemented on many websites. The Google search algorithm called Hummingbird update with greater dependence on Knowledge Graph has made schema.org more important. Google wants to categorize the web and specify specific entity points, schema.org is one way to make this happen. It is a collaboration project between *Google, Bing, Yahoo!* and *Yandex*. These search engines all support schema.org markup, which not necessarily ignore other types of markup such as microformats or RDFa. There must to be said that not all of the markup creates rich snippets or enhanced search results. "Rich snippets" refers to elements appearing in a SERP listing that are not a meta description. Google may display „Rich snippets" for many types of content such as events, music, organizations, people, products, recipes, reviews, software applications, videos and more.

The schema.org vocabulary is used along with the microdata format, to add information to HTML content. While the long term goal is to support a wider range of formats, the initial focus is on Microdata. According to [6], Microdata stands for machine-readable data that can be embedded in HTML documents in an easy-to-write manner, with an unambiguous parsing model. It is compatible with numerous other data formats including RDF [8] and JSON [7].

For implementation the schena.org presents the next example. This example is presented by comparison of using without and with the use of microformats.

First example is using html content without microformats:

```
<div>
 <h1>Avatar</h1>
 <span>Director: James Cameron (born August 16, 1954)</span>
 <span>Science fiction</span>
 <a href="../movies/avatar-theatrical-trailer.html">Trailer</a>
</div>
```

The second example uses the schema.org microformats standars:

```
<div itemscope itemtype ="http://schema.org/Movie">
  <h1 itemprop="name"&g;Avatar</h1>
  <div itemprop="director" itemscope itemtype="http://schema.org/Person">
  Director: <span itemprop="name">James Cameron</span> (born <span
itemprop="birthDate">August 16, 1954)</span>
  </div>
  <span itemprop="genre">Science fiction</span>
  <a href="../movies/avatar-theatrical-trailer.html"
itemprop="trailer">Trailer</a>
</div>
```

In the last example we must emphasize the schema.org elements. By adding **itemscope**, we are specifying that the HTML contained in the ... block is about a particular item. And the type of item can be specified using the **itemtype** attribute immediately after the itemscope. This specifies that the item contained in the div is in fact a Movie, as defined in the schema.org type hierarchy. Item types are provided as URLs, in this case http://schema.org/Movie. Movies have interesting properties such as actors, director, ratings. To label properties of an item, use the itemprop attribute. For example, to identify the director of a movie, add itemprop="director" to the element enclosing the director's name. Search engines can now understand not just that http://www.avatarmovie.com is a URL, but also that it's the URL for the trailer for the science-fiction movie Avatar, which was directed by James Cameron. Sometimes the value of an item property can itself be another item with its own set of properties. For example, we can specify that the director of the movie is an item of type Person and the Person has the properties name and birthDate. To specify that the value of a property is another item, you begin a newitemscope immediately after the corresponding itemprop.

The Open Graph protocol

A web site and in particular a web page must also have **S**ocial **M**edia **O**ptimizations implemented. This so called SMO's prepares the page to be shared on social media websites and needs some coding to be done. According to [9] the Open Graph protocol enables any web page to become a rich object in a social graph. It is used on Facebook to allow any web page to have the same functionality as any other object on Facebook. While many different technologies and schemas exist and could be combined together, there isn't a single technology which provides enough information to richly represent any web page within the social graph. To turn a web page into graph objects, we need to add basic metadata to the page. The Open Graph developers based the initial version of the protocol on RDFa which means that additional <meta> tags need to be placed in the <head> of the web page. The four required properties for every page are:

- og:title - The title of the object as it should appear within the graph;
- og:type - The type of the object, e.g., "video.movie". Depending on the type specified, other properties may also be required;
- og:image - An image URL which should represent the object within the graph;
- og:url - The canonical URL of the object that will be used as it is the permanent ID in the graph.

As an example, the following is the Open Graph protocol markup for The Rock on IMDB:

```
<html prefix="og: http://ogp.me/ns#">
<head>
<title>The Rock (1996)</title>
<meta property="og:title" content="The Rock" />
<meta property="og:type" content="video.movie" />
<meta property="og:url" content="http://www.imdb.com/title/tt0117500/" />
<meta property="og:image" content="http://ia.media-
imdb.com/images/rock.jpg" />
...
</head>
...
</html>
```

Another OG implementation example:

```
<meta property="og:locale" content="en_US" />
<meta property="og:type" content="article" />
<meta property="og:site_name" content="The Lightorialist" />
<meta property="og:url" content="http://lightorialist.com/the-story-of-
freedom-pictured-by-jake-olson/" />
<meta property="og:title" content="The story of freedom pictured by Jake
Olson" />
<meta property="og:image" content="http://lightorialist.com/wp-
content/uploads/userimages/gabi/the-story-of-freedom-pictured-by-jake-
olson/lightorialist-20141009-reagan-kays-of-omaha-nebraska-jake-olson-
studios-709x1024.jpg" />
<meta property="og:description" content="At first view, looking at Jake
Olson' gallery a strange envy reveal to stroke the prisoner of my own
narrowness. It's such a long time for me to remember how freedom feels
like ... How ..." />
<meta property="og:updated_time" content="2014-10-
09T11:13:13+00:00" />
<meta property="og:image" content="http://lightorialist.com/wp-
content/uploads/userimages/gabi/the-story-of-freedom-pictured-
by-jake-olson/lightorialist-20141009-miss-rebekah-romero-from-
harlan-iowa-jake-olson-studios-1024x737.jpg" />
```

3. STANDARD HTML TAGS

For highlighting and outlining the important phrases and keywords that characterizes the written web page we, as web developers, must do some programming that will then be used as a template for every page we display and promote online over the Internet.

The coding that involves the use of HTML tags needs to outline the following:

Title:

The title of a web page must be short, relevant and contain the most important keywords. The title of the page must be included in the html header section of the web page between the tags <title> and </title>. The inline title must be included between the <h1> and </h1>tags.

Description

The description of a web page must be short and relevant and included in the html header section in the tag like this: <meta name="description" content="a relevant description for SERP's">. This tag ensures that the displayed content in the SERP's is something we want to be displayed and not what the Search Engine decides by itself. This tag must be strategically filled with information so that the user is well informed about the content before clicking the reference link.

Keywords

The keywords must be added in the html header section of the web page and it's the standard way of telling Search Engines which are the most important and relevant words upon the web page is based on. The keywords are included in the html header section in the tag: <meta name="keywords" content="HTML, CSS, XML, XHTML, JavaScript">

Author

By providing each web page an author name the content gets more credibility not only for SEO but for visitors as well. The author meta tag must be included in the html header section: <meta name="author" content="*Name SName*">

Date

Each web page offers information on a subject and there are subjects that are time dependent. For such web pages we must declare the creation or publish date of the web page so that it can be identified on a time bases. The html header meta tag is: <meta name="date" content="2014-06-01T08:49:37+02:00">

Time

The html5 <time> tag defines either a time, or a date in the Gregorian calendar, optionally with a time and a time-zone offset. Example of using the inline time tag for attaching a temporal element to the presented web page content: <time datetime="2014-05-18T20:00+00:00">20pm article creation time </time>

4. XML AS A MACHINE UNDERSTANDABLE STANDARD FOR SHARING DATA

The main coding standard through which servers interact and exchange information is XML. This standard is used for servers to export database contents so that other can import it. Through this technology the exchange scope has diversified by implementing SEO uses such as Sitemaps for Search Engines and RSS Feeds for sharing contents to other web applications.

XML used by Sitemaps for Search Engines

Nowadays websites include many of own URL internal references in form of web links. For search engines can be difficult to understand the whole structure of a website and some guidelines must be offered that define the internal tree of links, taxonomies and categories. To accomplish this machine understandable guideline we implement sitemaps based on XML standard coding. The sitemaps for Google can be uploaded and managed through the webmaster tools. With this tool we can keep track of how many links we sent through the sitemap and how many have been indexed. Through the process of indexing can occur some errors that we can see and manage through the offered tool. According to [10] Sitemap is an XML file that lists URLs for a site along with additional metadata about each URL which specifies when it was last updated, how often it usually changes, and how important it is, relative to other URLs in the site so that search engines can more intelligently crawl the site. Here is some example of a Sitemap observed from the website http://lightorialist.com/post-sitemap.xml. The example bellow uses Sitemap 0.9 which is offered under the terms of the Attribution-Share Alike Creative Commons License and has wide adoption, including support from Google, Yahoo!, and Microsoft.

```
<?xml version="1.0" encoding="UTF-8"?><?xml-stylesheet type="text/xsl"
href="//lightorialist.com/main-sitemap.xsl"?>
<urlset xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xmlns:image="http://www.google.com/schemas/sitemap-image/1.1"
xsi:schemaLocation="http://www.sitemaps.org/schemas/sitemap/0.9
http://www.sitemaps.org/schemas/sitemap/0.9/sitemap.xsd"
xmlns="http://www.sitemaps.org/schemas/sitemap/0.9">
 <url>
  <loc>http://lightorialist.com</loc>
  <lastmod>2014-05-21T07:18:44+00:00</lastmod>
  <changefreq>daily</changefreq>
  <priority>1</priority>
 </url>
 <url>
  <loc>http://lightorialist.com/time-lapse-video-furious-clouds/</loc>
  <lastmod>2014-04-28T15:27:15+00:00</lastmod>
  <changefreq>weekly</changefreq>
  <priority>0.6</priority>
  <image:image>
   <image:loc>http://lightorialist.com/wp-
content/uploads/userimages/gabi/time-lapse-video-furious-clouds/time-
lapse-video-of-furious-clouds-8226.jpg</image:loc>
   <image:caption>IMG 9849</image:caption>
  </image:image>
 </url>
 <url>
  <loc>http://lightorialist.com/relaxing-ocean-waves-beaches-ocean-waves-
sound/</loc>
  <lastmod>2014-04-28T15:27:15+00:00</lastmod>
  <changefreq>weekly</changefreq>
  <priority>0.6</priority>
  <image:image>
```

```
   <image:loc>http://lightorialist.com/wp-
content/uploads/userimages/gabi/relaxing-ocean-waves-beaches-ocean-waves-
sound/lightorialist-image-ocean-waves-relaxation.jpg</image:loc>
   <image:caption>Ocean Waves relaxation video</image:caption>
  </image:image>
 </url>
</urlset>
```

According to [11] all data values in a Sitemap must be entity-escaped. The generated XML file must be UTF-8 encoded. The Sitemap must:

- Begin with an opening <urlset> tag and end with a closing </urlset> tag.
- Specify the namespace (protocol standard) within the <urlset> tag.
- Include a <url> entry for each URL, as a parent XML tag.
- Include a <loc> child entry for each <url> parent tag.

XML for RSS feed

According to [12] RSS has its name origins from RDF Site Summary and is often dubbed with Really Simple Syndication. RSS uses standard web feed formats to publish frequently updated information: news, blog, audio, and video. An RSS document is also known as "feed", "web feed" or "channel". The RSS includes full or summarized text, and metadata, like publishing date and author's name. RSS feeds enable publishers to syndicate data automatically.

A start example of XML tags used for RSS feed taken from the website http://www.lightorialist.com is shown below and presents the main tags such as *rss, channel, title, item, link*:

```
<rss xmlns:content="http://purl.org/rss/1.0/modules/content/" xmlns:wfw="
http://wellformedweb.org/CommentAPI/" xmlns:dc="http://purl.org/dc/elemen
ts/1.1/" xmlns:atom="http://www.w3.org/2005/Atom"xmlns:sy="http://purl.or
g/rss/1.0/modules/syndication/" xmlns:slash="http://purl.org/rss/1.0/modu
les/slash/" version="2.0">
<channel>
<title>The Lightorialist</title>
<atom:link href="http://lightorialist.com/feed/" rel="self" type="applica
tion/rss+xml"/>
<link>http://lightorialist.com</link>
<description>What we see, we become!</description>
<lastBuildDate>Thu, 15 May 2014 14:10:20 +0000</lastBuildDate>
<language>en-US</language>
<sy:updatePeriod>hourly</sy:updatePeriod>
<sy:updateFrequency>1</sy:updateFrequency>
<item>
<title>Buddha's Birthday Lanterns Shots</title>
<link>
http://lightorialist.com/buddhas-birthday-lanterns/
</link>
</item>
</channel>
</rss>
```

A standard XML file format ensures compatibility with many different machines and applications. RSS feeds also benefit users who want to receive timely updates from favorite websites or to aggregate data from many sites. Subscribing to a website RSS removes the need for the user to manually check the web site for new content. Instead, their browser constantly monitors the site and informs the user of any updates. The browser can also be commanded to automatically download the new data for the user.

5. CONCLUSION

State of the art SEO codding will always get much more relevance and credibility not only by Search Engines but also by the actual visitor of the web page who will give more credit to the content and generates more click through and a lowered bouncerate. The guidelines presented in this article I found to be the most important in terms of coding. In general terms of SEO there are also other very important actions to get involved with and taken care of such as: backlinks; authorship; speed of source webpage loading and automated or manually Social Media Sharing of relevant website content through targeted so called Social Media Groups and Communities that are relevant for the subject you want to share. Writing, targeting and sharing relevant web content on specific subjects is the most important step in promoting a reliable and Non-Spam Web environment for all of us.

6. REFERENCES:

[1] http://dublincore.org/documents/dces/
[2]http://www.quicksprout.com/2012/01/06/6-on-page-seo-strategies-that-will-boost-your-rankings/
[3] http://blog.woorank.com/2013/04/dublin-core-metadata-for-seo-and-usability/
[4] https://schema.org/
[5]http://searchengineland.com/schema-org-7-things-for-seos-to-consider-post-hummingbird-172163
[6] http://dev.w3.org/html5/md-LC/
[7] http://www.json.org/
[8] http://www.w3.org/RDF/
[9] http://ogp.me/
[10] http://www.sitemaps.org/
[11] http://www.sitemaps.org/protocol.html
[12] http://en.wikipedia.org/wiki/RSS
[13] http://lightorialist.com/the-story-of-freedom-pictured-by-jake-olson/
[14] Pîrjan Alexandru, Optimization Techniques for Data Sorting Algorithms, The 22nd International DAAAM Symposium, Vienna, 2011, Annals of DAAAM for 2011 & Proceedings of the 22nd International DAAAM Symposium, Vienna, 2011, pp. 1065-1066, ISSN 1726-9679, ISBN 978-3-901509-73-5.