Abstract

This memo describes a method of registering HTML documents with a specific geographic location through means of embedded META tags. The content of the META tags gives the geographic position of the resource described by the HTML document in terms of Latitude, Longitude, and optionally Elevation in a simple, machine-readable manner. This information may be used for automated resource discovery by means of an HTML indexing agent or search engine. META tags giving a civic location of a resource are also described.

1. Introduction

Many resources described by HTML documents on the World-Wide-Web are
associated with a particular place on the Earth’s surface. While resource discovery on the Web has thus far focussed on document title and open-text keyword searching, in these cases it may be beneficial to facilitate geographic searching. Examples of this kind of resource include pages describing restaurants, shipwrecks, retail stores etc. Consumers may use this information in order to select the closest facility, and in order to navigate towards a resource by road, on foot or by other means.

Although some resources, such as restaurants, have a street address which may be mapped to geographic location by existing tools, other objects on the Web, such as a photograph of a mountain, may not.

This draft describes a method of adding static location data to legacy HTML documents using a construct that is familiar to many HTML authors. It is intended to be concise, unambiguous, simple to use and compatible with existing editing tools. The intended use is to provide location data to Web robots that typically revisit pages every few weeks.

It is anticipated that in many cases this location data will be added manually by persons unfamiliar with GIS terminology or metadata standards. For this reason a minimal data set with few options is preferred over a more complex and extensible one.

The method described in this draft is not intended to preempt existing or future metadata encapsulation schemes which may better serve the needs of a particular community, such as geographic information systems (GIS). Nor is it intended to preempt richer, more structured data encapsulations such as RDF or XML, which typically require software to generate correctly.

2. Coordinate Systems

Resource positions on the Earth’s surface should be expressed in degrees North of Latitude, degrees East of Longitude as signed decimal numbers.

Where the precision of the coordinates is such that the datum used is significant, typically more precise than one kilometre distance, positions should be converted to the WGS 84 datum [WGS84]. Elevations, if given, should be in metres above datum. Positions given by a GPS set [GPS] with datum set to "WGS 84" will in most cases be adequate, of the order of 15 metres accuracy in horizontal position and 25 metres in elevation.

It should be noted that elevations referred to the WGS 84 geoid will in some areas differ appreciably from those measured with respect to
local datum in coastal regions, which may be Mean High Water Springs, Mean Sea Level, Higher High Water or a similar reference level, and will differ substantially from "ground level". Use of elevation is not recommended unless its value may be reliably determined.

3. Implementation

XHTML, HTML or WML markup should be added to the document in the form of a META statement. This should be placed in the document head in accordance with the XHTML specification [XHTML].

There are several possible GEO identifiers:
The identifier "geo.position" is used for Latitude, Longitude and optionally Elevation data.
The identifier "geo.country" is used for the two-letter country code from ISO 3166 [ISO3166], e.g. "US" (United States), "DE" (Germany).
The identifiers "geo.a1", "geo.a2" etc. are used to define a civic address, as in RFC 4776 [RFC4776].

For resources within the United States and Canada, the "geo.a1" identifier corresponds to and the common 2-character State/Province codes [STATES][PROVINCES], e.g. "BC" (British Columbia), "CA" (California).

To facilitate machine indexing, wherever possible a controlled list should be used for civic elements. For instance, ISO 3166-2 [ISO3166-2] might be used for "geo.a1".

Use of the numeric "geo.position" is generally recommended to ensure accurate indexing. However, if the resource described is localized to a country or region, but not to a single point, the civic identifiers "geo.country", "geo.a1" etc. may be used alone without a corresponding "geo.position" identifier.

It is the intention of this draft to provide a means to associate a single point with an XHTML, HTML or WML document. Some consideration should be given to the choice of location when describing a resource, given that positioning mechanisms may provide an accuracy of the order of ten metres in horizontal position. For instance, when describing a retail store or small business, it may be more meaningful to give the position of the street entrance rather than the position of the center of the property.

Although the XHTML specification [XHTML] states that the name field is in general case-sensitive, these GEO tags should be recognized by
compliant agents regardless of case. Coordinates should be ordered (Latitude ; Longitude) as for RFC 2426, RFC 2445 (vCard and iCal specifications) [ICAL][VCARD]. If elevation is given, coordinates should be ordered (Latitude ; Longitude ; Elevation).

3.1 Migration from earlier versions

To migrate documents and applicaitons written against earlier versions of this draft, the following correspondences are noted:

    geo.position       geo.position
    geo.region         geo.country   (2 character region)
    geo.region         geo.country and geo.al (extended region XX-YYY)
    geo.placename      geo.lmk     (landmark or vanity address)

4. Examples

    <meta name="geo.position" content="48.54;-123.84;115" />

describes a resource 115 metres above datum at position 48.54 degrees North, 123.84 degrees West, while

    <meta name="geo.position" content="-10;60" />

describes a resource at position 10 degrees South, 60 degrees East.

    <meta name="geo.country" content="CA" />
    <meta name="geo.al" content="ON" />
    <meta name="geo.a3" content="London" />

describes a resource in London, Ontario, Canada, while

    <meta name="geo.country" content="GB" />
    <meta name="geo.al" "LND" />
    <meta name="geo.a3" content="London" />

describes a resource in London, England (Great Britain).

5. Semantics

Values for latitude and longitude shall be expressed as decimal fractions of degrees. Whole degrees of latitude shall be represented by a decimal number ranging from 0 through 90. Whole degrees of longitude shall be represented by a decimal number ranging from 0 through 180. When a decimal fraction of a degree is specified, it shall be separated from the whole number of degrees by a decimal
point (the period character, "."). Decimal fractions of a degree should be expressed to the precision available, with trailing zeroes being used as placeholders if required. A decimal point is optional where the precision is less than one degree. Some effort should be made to preserve the apparent precision when converting from another datum or representation, for example 41 degrees 13 minutes should be represented as 41.22 and not 41.21666, while 41 13’ 11” may be represented as 41.2197.

Latitudes north of the equator MAY be specified by a plus sign (+), or by the absence of a minus sign (-), preceding the designating degrees. Latitudes south of the Equator MUST be designated by a minus sign (-) preceding the digits designating degrees. Latitudes on the Equator MUST be designated by a latitude value of 0.

Longitudes east of the prime meridian shall be specified by a plus sign (+), or by the absence of a minus sign (-), preceding the designating degrees. Longitudes west of the prime meridian MUST be designated by a minus sign (-) preceding the digits designating degrees. Longitudes on the prime meridian MUST be designated by a longitude value of 0. A point on the 180th meridian shall be taken as 180 degrees West, and shall include a minus sign.

Any spatial address with a latitude of +90 (90) or -90 degrees will specify a position at the True North or True South Poles, respectively. The component for longitude may have any legal value.

The vertical coordinate (Elevation) must be expressed in meters above WGS-84 (EGM96) datum. Points having zero elevation must not have a negative sign.

5.1 Interpretation

User agents should accept metadata written according to the HTML or XHTML specifications [HTML][XHTML].

Whitespace within a position value shall be ignored.

An interpreting agent shall internally mark position values either valid or invalid. If a position is marked invalid, it shall not be used to index or qualify the containing document.

A position having a Latitude greater than 90 degrees, or less than -90 degrees, shall be marked invalid.
A position having a Longitude greater than 180 degrees, or less than
-180 degrees, shall be marked invalid.

Where a value is given for geo.country, and the latitude and
longitude values given for geo.position fall outside the recognized
boundaries of this region, the position may be marked invalid. For
example, if a country code of "US" is given for a location in the US
mainland, the position may be marked invalid if the Latitude is
negative or the Longitude is positive.

No formal reliance shall be placed on the precision implicit in
position data. It is likely that few content providers are qualified
to determine reliable precision or accuracy data, and may use
position data from other sources which does not give the datum.

5.2. Terminology

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT",
"SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this
document are to be interpreted as described in [RFC2119].

6. Formal Syntax

DIGIT = %x30-39 ; 0-9
PLUS = %x2B ; +
MINUS = %x2D ; -
DECIMAL = %x2E ;
SEMI = %x3B ; ;
CRLF = %x0D.%x0A ; return, linefeed
SP = %x20 ; space
HTAB = %x09 ; tab
WSP = SP / HTAB ;
LWSP = (WSP / CRLF WSP) ; linear whitespace
UCASE = %x41-5A ; A-Z
HYPHEN = %x2D ; -
USCORE = %x5F ; _
country = 2UCASE ; 2-letter code from ISO3166
TEXT = <any OCTET except CTLs, but including LWSP>
placename = 1*TEXT
delimiter = SEMI

latitude = [ MINUS / PLUS ] 0*2DIGIT [ DECIMAL *DIGIT]
longitude = [ MINUS / PLUS ] 0*3DIGIT [ DECIMAL *DIGIT]
elevation = [ MINUS / PLUS ] 0*DIGIT [ DECIMAL *DIGIT]
position = latitude <delimiter> longitude [ <delimiter> elevation ]

geocivic = TEXT ; civic address elements as per RFC 4776
XHTML or WML syntax:
<meta name="geo.position" content="<position>" />
<meta name="geo.country" content="<country>" />
<meta name="geo.<civic>" content="<TEXT>" />

HTML (legacy) syntax:
<meta name="geo.position" content="<position>">
<meta name="geo.country" content="<country>">
<meta name="geo.<civic>" content="<TEXT>">

7. Applicability

As stated in the introduction, certain HTML documents may be associated with a geographic position, while other documents are not. For proper use of the GEO tags as described in this draft, the resource described in an HTML document should be associated with a particular geographic location for the lifetime of the document. The tags may thus be properly used to describe an object fixed on the surface of the earth (or more properly, fixed in position relative to the surface of the earth) such as a retail store, a mountain peak or a railway station. They may not be used to describe a non-localized, moving, or intangible object such as a multinational company, river, aircraft or mathematical theory.

The geographic position given is associated with the resource described by the HTML document, not with the physical location of the document [RFC1876], or the location of the company responsible for publishing or hosting the document. Thus, in some cases the country code used in "geo.country" may differ from the country code forming part of the host address in the document URL.

Since the position given is associated with the content of the document, not the author, publishing and document conversion tools should not cache position data or store it in a template.

In cases where the object being described is an area, such as a lake or a building, the position of the object should not in general be given to greater precision than the width of the object. If desired, features within the object may be described in another page and their position given with greater precision. In the case of an object such as a place of business, where only one page exists, the position of the entrance may be given rather than the position of the centroid.

8. Security Considerations
The intended use of GEO metadata as described in this draft raises no privacy issues beyond those associated with normal use of the Web. It is assumed that information present in public Web pages has been published in accordance with applicable privacy regulations and guidelines.

If the location data describes the position of a mobile Internet device, filters applicable to possible end recipients (typically, the public Internet) should be applied. The webserver in this case acts as a Location Recipient [RFC3693].

9. Internationalization considerations

HTML meta element content, including geo elements, is coded using the character set of the containing document, typically UTF-8 or ISO8859-1.

Geo.country and geo.position tag content should contain only ASCII characters.

10.1 Normative References


10.2 Informative References

http://www.ietf.org/rfc/rfc1876.txt

http://www.ietf.org/rfc/rfc4776.txt


[ICAL] Dawson & Stenerson, Internet Calendaring and Scheduling Core Object Specification (iCalendar), RFC 2445, November 1998
http://www.ietf.org/rfc/rfc2445.txt

http://www.ietf.org/rfc/rfc2426.txt

[STATES] United States Postal Service, Official Abbreviations - States and Possessions,
http://www.usps.gov/ncsc/lookups/abbr_state.txt

[PROVINCES] Canada Postal Guide, Province and Territory Symbols
http://www.canadapost.ca/tools/pg/manual/b03-e.asp

11. Acknowledgments
Rohan Mahy and Patrik Faltstrom of Cisco Systems, for semantics.

12. Authors’ Addresses

Andrew Daviel, BSc.
Vancouver Webpages, Box 357
Intellectual Property Statement
The IETF takes no position regarding the validity or scope of any Intellectual Property Rights or other rights that might be claimed to pertain to the implementation or use of the technology described in this document or the extent to which any license under such rights might or might not be available; nor does it represent that it has made any independent effort to identify any such rights. Information on the procedures with respect to rights in RFC documents can be found in BCP 78 and BCP 79.

Copies of IPR disclosures made to the IETF Secretariat and any assurances of licenses to be made available, or the result of an attempt made to obtain a general license or permission for the use of such proprietary rights by implementers or users of this specification can be obtained from the IETF on-line IPR repository at http://www.ietf.org/ipr.

The IETF invites any interested party to bring to its attention any copyrights, patents or patent applications, or other proprietary rights that may cover technology that may be required to implement this standard. Please address the information to the IETF at ietf-ipr@ietf.org.

Disclaimer of Validity
This document and the information contained herein are provided on an "AS IS" basis and THE CONTRIBUTOR, THE ORGANIZATION HE/SHE REPRESENTS OR IS SPONSORED BY (IF ANY), THE INTERNET SOCIETY, THE IETF TRUST AND THE INTERNET ENGINEERING TASK FORCE DISCLAIM ALL WARRANTIES, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO ANY WARRANTY THAT THE USE OF THE INFORMATION HEREIN WILL NOT INFRINGE ANY RIGHTS OR ANY IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE.

Copyright Statement
Copyright (C) The IETF Trust (2007).

This document is subject to the rights, licenses and restrictions contained in BCP 78, and except as set forth therein, the authors retain all their rights.

Acknowledgment
Funding for the RFC Editor function is currently provided by the Internet Society.

15a. IANA Considerations
This document does not introduce any IANA considerations.