Network Working Group
Request for Comments: 3236
Category: Informational

M. Baker
Planetfred, Inc.
P. Stark
Ericsson Mobile Communications

January 2002

The 'application/xhtml+xml' Media Type

Status of this Memo

Copyright Notice

Abstract

This document defines the 'application/xhtml+xml' MIME media type for XHTML based markup languages; it is not intended to obsolete any previous IETF documents, in particular RFC 2854 which registers 'text/html'.

1. Introduction

In 1998, the W3C HTML working group began work on reformulating HTML in terms of XML 1.0 [XML] and XML Namespaces [XMLNS]. The first part of that work concluded in January 2000 with the publication of the XHTML 1.0 Recommendation [XHTML1], the reformulation for HTML 4.01 [HTML401].

Work continues in the Modularization of XHTML Recommendation [XHTMLM12N], the decomposition of XHTML 1.0 into modules that can be used to compose new XHTML based languages, plus a framework for supporting this composition.

This document only registers a new MIME media type, 'application/xhtml+xml'. It does not define anything more than is required to perform this registration.

This document follows the convention set out in [XMLMIME] for the MIME subtype name; attaching the suffix "+xml" to denote that the entity being described conforms to the XML syntax as defined in XML 1.0 [XML].

This document was prepared by members of the W3C HTML working group based on the structure, and some of the content, of RFC 2854, the registration of 'text/html'. Please send comments to www-html@w3.org, a public mailing list (requiring subscription) with archives at <http://lists.w3.org/Archives/Public/www-html/>.

2. Registration of MIME media type application/xhtml+xml

MIME media type name: application
MIME subtype name: xhtml+xml
Required parameters: none
Optional parameters:

charset
This parameter has identical semantics to the charset parameter of the "application/xml" media type as specified in [XMLMIME].

profile
See Section 8 of this document.

Encoding considerations:
See Section 4 of this document.

Security considerations:
See Section 7 of this document.

Interoperability considerations:
XHTML 1.0 [XHTML10] specifies user agent conformance rules that dictate behaviour that must be followed when dealing with, among other things, unrecognized elements.

With respect to XHTML Modularization [XHTMLMOD] and the existence of XHTML based languages (referred to as XHTML family members) that are not XHTML 1.0 conformant languages, it is possible that 'application/xhtml+xml' may be used to describe some of these documents. However, it should suffice for now for the purposes of interoperability that user agents accepting 'application/xhtml+xml' content use the user agent conformance rules in [XHTML1].

Although conformant 'application/xhtml+xml' interpreters can expect that content received is well-formed XML (as defined in [XML]), it cannot be guaranteed that the content is valid XHTML (as defined in [XHTML1]). This is in large part due to the reasons in the preceding paragraph.

Published specification:
XHTML 1.0 is now defined by W3C Recommendation; the latest published version is [XHTML1]. It provides for the description of some types of conformant content as "text/html", but also doesn't disallow the use with other content types (effectively allowing for the possibility of this new type).

Applications which use this media type:
Some content authors have already begun hand and tool authoring on the Web with XHTML 1.0. However that content is currently described as "text/html", allowing existing Web browsers to process it without reconfiguration for a new media type.

There is no experimental, vendor specific, or personal tree predecessor to 'application/xhtml+xml'. This new type is being registered in order to allow for the expected deployment of XHTML on the World Wide Web, as a first class XML application where authors can expect that user agents are conformant XML 1.0 [XML] processors.

Additional information:

Magic number:
There is no single initial byte sequence that is always present for XHTML files. However, Section 5 below gives some guidelines for recognizing XHTML files. See also section 3.1 in [XMLMIME].

File extension:
There are three known file extensions that are currently in use for XHTML 1.0; ".xht", ".xhtml", and ".html".

It is not recommended that the ".xml" extension (defined in [XMLMIME]) be used, as web servers may be configured to distribute such content as type "text/xml" or "application/xml". [XMLMIME] discusses the unreliability of this approach in section 3. Of course, should the author desire this behaviour, then the ".xml" extension can be used.

Macintosh File Type code: TEXT

Person & email address to contact for further information:
Mark Baker <mark.baker@canada.sun.com>

Intended usage: COMMON

Author/Change controller:
The XHTML specifications are a work product of the World Wide Web Consortium's HTML Working Group. The W3C has change control over these specifications.

3. Fragment identifiers

URI references (Uniform Resource Identifiers, see [RFC2396] as updated by [RFC2732]) may contain additional reference information, identifying a certain portion of the resource. These URI references end with a number sign ("#") followed by an identifier for this portion (called the "fragment identifier"). Interpretation of fragment identifiers is dependent on the media type of the retrieval result.

For documents labeled as 'text/html', [RFC2854] specified that the fragment identifier designates the correspondingly named element, these were identified by either a unique id attribute or a name attribute for some elements. For documents described with the application/xhtml+xml media type, fragment identifiers share the same syntax and semantics with other XML documents, see [XMLMIME], section 5.

At the time of writing, [XMLMIME] does not define syntax and semantics of fragment identifiers, but refers to "XML Pointer Language (XPointer)" for a future XML fragment identification mechanism. The current specification for XPointer is available at http://www.w3.org/TR/xptr. Until [XMLMIME] gets updated, fragment identifiers for XHTML documents designate the element with the corresponding ID attribute value (see [XML] section 3.3.1); any XHTML element with the "id" attribute.

4. Encoding considerations

By virtue of XHTML content being XML, it has the same considerations when sent as 'application/xhtml+xml' as does XML. See [XMLMIME], section 3.2.

5. Recognizing XHTML files

All XHTML documents will have the string "<html" near the beginning of the document. Some will also begin with an XML declaration which begins with "<?xml", though that alone does not indicate an XHTML document. All conforming XHTML 1.0 documents will include an XML document type declaration with the root element type 'html'.

XHTML Modularization provides a naming convention by which a public identifier for an external subset in the document type declaration of a conforming document will contain the string "//DTD XHTML". And while some XHTML based languages require the doctype declaration to occur within documents of that type, such as XHTML 1.0, or XHTML Basic (http://www.w3.org/TR/xhtml-basic), it is not the case that all XHTML based languages will include it.

All XHTML files should also include a declaration of the XHTML namespace. This should appear shortly after the string "<html", and should read 'xmlns="http://www.w3.org/1999/xhtml"'.

6. Charset default rules

By virtue of all XHTML content being XML, it has the same considerations when sent as 'application/xhtml+xml' as does XML. See [XMLMIME], section 3.2.

7. Security Considerations

The considerations for "text/html" as specified in [TEXTHTML] and and for 'application/xml' as specified in [XMLMIME], also hold for 'application/xhtml+xml'.

In addition, because of the extensibility features for XHTML as provided by XHTML Modularization, it is possible that 'application/xhtml+xml' may describe content that has security implications beyond those described here. However, if the user agent follows the user agent conformance rules in [XHTML1], this content will be ignored. Only in the case where the user agent recognizes and processes the additional content, or where further processing of that content is dispatched to other processors, would security issues potentially arise. And in that case, they would fall outside the domain of this registration document.

8. The "profile" optional parameter

This parameter is meant to solve the short-term problem of using MIME media type based content negotiation (such as that done with the HTTP "Accept" header) to negotiate for a variety of XHTML based languages. It is intended to be used only during content negotiation. It is not expected that it be used to deliver content, or that origin web servers have any knowledge of it (though they are welcome to). It is primarily targeted for use on the network by proxies in the HTTP chain that manipulate data formats (such as transcoders).

The parameter is intended to closely match the semantics of the "profile" attribute of the HEAD element as defined in [HTML401] (section 7.4.4.3), except it is applied to the document as a whole rather than just the META elements. More specifically, the value of the profile attribute is a URI that can be used as a name to identify a language. Though the URI need not be resolved in order to be useful as a name, it could be a namespace, schema, or a language specification.

As an example, user agents supporting only XHTML Basic (see http://www.w3.org/TR/xhtml-basic) currently have no standard means to convey their inability to support the additional functionality in XHTML 1.0 [XHTML1] that is not found in XHTML Basic. While XHTML Basic user agent conformance rules (which are identical to XHTML 1.0) provide some guidance to its user agent implementators for handling some additional content, the additional content in XHTML 1.0 that is not part of XHTML Basic is substantial, making those conformance rules insufficient for practical processing and rendering to the end user. There is also the matter of the potentially substantial burden on the user agent in receiving and parsing this additional content.

The functionality afforded by this parameter can also be achieved with at least two other more general content description frameworks; the "Content-features" MIME header described in RFC 2912, and UAPROF from the WAPforum (see http://www.wapforum.org/what/technical.htm). At this time, choosing one of these solutions would require excluding the other, as interoperability between the two has not been defined. For this reason, it is suggested that this parameter be used until such time as that issue has been addressed.

An example use of this parameter as part of a HTTP GET transaction would be;

```
Accept: application/xhtml+xml;
  profile="http://www.w3.org/TR/xhtml-basic/xhtml-basic10.dtd"
```

9. Author's Address

Mark A. Baker
Planetfred, Inc.
44 Byward Market, Suite 240
Ottawa, Ontario, CANADA. K1N 7A2
Phone: +1-613-789-1818
EMail: mbaker@planetfred.com
EMail: distobj@acm.org

Peter Stark
Ericsson Mobile Communications
Phone: +464-619-3000
EMail: Peter.Stark@ecs.ericsson.com

10. References

[HTML401] Raggett, D., et al., "HTML 4.01 Specification", W3C Recommendation. Available at <http://www.w3.org/TR/html401> (or <http://www.w3.org/TR/1999/REC-html401-19991224>).

[MIME] Freed, N. and N. Borenstein, "Multipurpose Internet Mail Extensions (MIME) Part Two: Media Types", RFC 2046, November 1996.

[URI] Berners-Lee, T., Fielding, R. and L. Masinter, "Uniform Resource Identifiers (URI): Generic Syntax", RFC 2396, August 1998.

[XHTML1] "XHTML 1.0: The Extensible HyperText Markup Language: A Reformulation of HTML 4 in XML 1.0", W3C Recommendation. Available at <http://www.w3.org/TR/xhtml1>.

[XML] "Extensible Markup Language (XML) 1.0", W3C Recommendation. Available at <http://www.w3.org/TR/REC-xml> (or <http://www.w3.org/TR/2000/REC-xml-20001006>).

[TEXTHTML] Connolly, D. and L. Masinter, "The 'text/html' Media Type", RFC 2854, June 2000.

[XMLMIME] Murata, M., St.Laurent, S. and D. Kohn, "XML Media Types", RFC 3023, January 2001.

[XHTMLM12N] "Modularization of XHTML", W3C Recommendation. Available at: <http://www.w3.org/TR/xhtml-modularization>

11. Full Copyright Statement

Acknowledgement

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