Change Request for registered IPTC MIME types

Author: Misha Wolf, Reuters
Date: 2003-09-12

Change the NewsML MIME type to application/vnd.IPTC.NewsML+xml
and the NITF MIME type to application/vnd.IPTC.NITF+xml

The IPTC has registered the MIME types:
text/vnd.IPTC.NewsML
text/vnd.IPTC.NITF

These have all the disadvantages of not being identified as XML, as listed in RFC 3023 ("XML Media Types"):
http://www.ietf.org/rfc/rfc3023.txt).

We propose the registration of:
application/vnd.IPTC.NewsML+xml
application/vnd.IPTC.NITF+xml
followed by official deprecation of the old MIME types.

As background, there follows an extract from a mail Rob Warner, a Reuters Technical Director, sent to the newsml@yahoogroups.com list, and some extracts from RFC 3023.

<RobWarner>

In the case of NewsML, the fact that it's based on XML means the content has a well-defined structure which, in an of itself, has utility, separate from the actual meaning of the data.

This fact allows one to use powerful technologies such as XPointer & XLink to refer to elements within arbitrary XML documents, without having to understand the actual grammar in use. It also allows search engines to do much more flexible things with the data than if it were treated as either an opaque chunk of binary, or just plain text.

Of course it's true that the type can be determined by inspection, but in an HTTP context that would require retrieving the whole document, and we all know that NewsML is not a terse grammar, so that would be an expensive operation.

If we have a suitably expressive MIME type, full body retrieval wouldn't be necessary - the headers can be retrieved separately, potentially saving a huge amount of extra bandwidth & processing.

Basically, if we don't have a MIME type ending in +xml we're effectively saying that XPointer, XLink & any other technology that exploits the fact that a document is XML cannot be used efficiently on NewsML, especially over slow links. That may limit the adoption of NewsML in certain circumstances (some developing countries for instance). Read §7 of http://www.ietf.org/rfc/ rfc3023.txt for more good reasons we should seriously consider assigning a new type.

Quite apart from the technical arguments, there's the "everyone else is doing it" factor - a quick search picked up the following types that are already assigned:

application/beep+xml
application/cnrp+xml
application/cpl+xml
application/reginfo+xml
application/vnd.criticaltools.wbs+xml
application/vnd.irepository.package+xml
application/vnd.liberty-request+xml
application/vnd.llamagraphics.life-balance.exchange+xml
application/vnd.mozilla.xul+xml
application/vnd.pwg-xhtml-print+xml
application/watcherinfo+xml
application/xhtml+xml

Some other, as-yet unregistered ones include image/svg+xml, application/sbml+xml, application/xenc+xml and application/mathml+xml.

Clearly all these groups see value in identifying their content additionally as XML, not just some opaque datatype that can only be understood by special tools. I feel strongly that it makes sense for NewsML to follow suit.

</RobWarner>

Extracts from RFC 3023 ("XML Media Types") follow. The third extract makes the most convincing argument for "application/*" as opposed to "text/*".

<RFC3023>

If an XML document -- that is, the unprocessed, source XML document -- is readable by casual users, text/xml is preferable to application/xml. MIME user agents (and web user agents) that do not have explicit support for text/xml will treat it as text/plain, for example, by displaying the XML MIME entity as plain text. Application/xml is preferable when the XML MIME entity is unreadable by casual users.

[...]

NOTE: Users are in general not used to text containing tags such as <price>, and often find such tags quite disorienting or annoying. If one is not sure, the conservative principle would suggest using application/* instead of text/* so as not to put information in front of users that they will quite likely not understand.

[...]

The top-level media type "text" has some restrictions on MIME entities and they are described in [RFC2045] and [RFC2046]. In particular, the UTF-16 family, UCS-4, and UTF-32 are not allowed (except over HTTP[RFC2616], which uses a MIME-like mechanism). Thus, if an XML document or external parsed entity is encoded in such

character encoding schemes, it cannot be labeled as text/xml or text/xml-external-parsed-entity (except for HTTP).

</RFC3023>

=== END of document ===