UDDI Spec TC

Technical Note

Handling of anyURI datatypes

Document identifier:

uddi-spec-tc-tn-anyurihandling

This version:

http://www.oasis-open.org/committees/uddi-spec/doc/tn/uddi-spec-tc-tn-anyurihandling-20040921.htm

Latest version:

http://www.oasis-open.org/committees/uddi-spec/doc/tn/uddi-spec-tc-tn-anyurihandling.htm

Author:

Claus von Riegen, SAP (claus.von.riegen@sap.com)

Editors:

Andrew Hately, IBM (hately@us.ibm.com)
Tony Rogers, CA (tony.rogers@ca.com)

Contributors:

Maud Cahuzac, France Télécom
Luc Clément, Systinet
John Colgrave, IBM
Matthew Dovey, Oxford University
Massimo Paolucci, CMU
Katia Sycara, CMU

Abstract:

Non-ASCII characters are supported by the XML Schema anyURI datatype but are not always supported in Web service tooling. This technical note describes the interoperability considerations when using anyURI-based data types in UDDI V3 API calls.

Status:

This document is updated periodically on no particular schedule. Send comments to the editor.

Committee members should send comments on this technical note to the uddi-spec@lists.oasis-open.org list. Others should subscribe to and send comments to the uddi-spec-comment@lists.oasis-open.org list. To subscribe, send an email message to uddi-spec-comment-request@lists.oasis-open.org with the word "subscribe" as the body of the message.

For information on whether any intellectual property claims have been disclosed that may be essential to implementing this technical note, and any offers of patent licensing terms, please refer to the Intellectual Property Rights section of the UDDI Spec TC web page (http://www.oasis-open.org/committees/uddi-spec/).

Table of Contents

1 Introduction

Non-ASCII characters are supported by the XML Schema anyURI datatype but are not always supported in Web service tooling. This technical note describes the interoperability considerations when using anyURI-based data types in UDDI V3 API calls.

1.1 Problem statement

While the set of characters allowed in legal URIs, as defined in **[RFC2396]** and amended in **[RFC2732]**, is restricted to ASCII characters, the set of characters allowed in the XML Schema datatype anyURI (see **[XMLSchema]**) is more flexible in that it actually allows most Unicode characters.

Whenever an XML element or attribute of type anyURI is to be converted to an actual URI, for example, to resolve a URL by using a Web browser, the algorithm defined in section 5.4 of **[XLink]** is to be used. This algorithm describes how non-ASCII characters are escaped to result in character sequences allowed in URIs as defined in RFC2732.

This support for the full character set allowed in anyURI is not implemented[1] in all Web Service or XML Schema aware validation tools.

Although this is not an issue unique to UDDI, the fact that UDDI Version 3
   a) defines several UDDI attributes (uddiKey) and elements (discoveryURL and overviewURL) to be of type anyURI
   b) requires the XML documents contained in a UDDI API request message to be validated against the respective UDDI XML Schema(s)
   c) requires UDDI nodes to preserve the content of such XML documents for subsequent use in UDDI API response messages
requires the use of XML tools. As a result, the use of multiple XML and Web Service tools or class libraries may result in different or incompatible behavior with regard to the handling of anyURI-based XML elements and attributes.

1.2 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]**.

---

[1] Discussions within the OASIS UDDI TC have shown that different TC members had different interpretations of what the allowed character set for anyURI based XML elements and attributes is. Clarifications with W3C representatives have shown that the allowed character set is unambiguously defined, as discussed in this Technical Note.

2 Technical note Solution

2.1 Technical note behavior

It is not possible to avoid the problem that XML tools in the past have been implemented differently in regard to handling anyURI-based elements and attributes. Some guidelines may help users and implementers deal appropriately where such tools are part of a UDDI node or client implementation.

UDDI nodes themselves may not be able to accept non-ASCII characters. In a multi-node UDDI registry, the registry MUST specify an "exception policy"[2] if any node in the registry can not handle Unicode anyURI elements. The character set allowed in anyURI data MUST be consistent across all nodes. In the case where registries do not accept non-ASCII characters publishers MUST encode the data in the anyURI element to produce an RFC2732 URI in UDDI Publication API calls.

Clients that can't handle the full Unicode character set don't represent a problem when publishing to the UDDI node. These clients only use a subset of the allowed characters in their UDDI API calls. However, clients themselves need to be prepared to receive non-ASCII characters in UDDI API response messages. It is also the responsibility of the client to apply the encoding and escaping [XLink] mechanism whenever a discoveryURL or overviewURL is to be resolved. In the case that a publisher is aware that some client software querying the registry may not handle Unicode anyURI elements, it is RECOMMENDED that the publisher encode the data in the anyURI element to produce an RFC2732 URI.

---

[2] Since this behavior is a deviation from the actual normative behavior of the UDDI V3 Specification and XML specifications, it is different from all other UDDI registry or node policies that simply specify how something is implemented in accordance to the specification.

3 References

3.1 Normative

[RFC2119]   S. Bradner, *Key words for use in RFCs to Indicate Requirement Levels*, http://www.ietf.org/rfc/rfc2119.txt, IETF RFC 2119, March 1997.

[RFC2396]   T. Berners-Lee, R. Fielding, L. Masinter, *Uniform Resource Identifiers (URI): Generic Syntax*, http://www.ietf.org/rfc/rfc2396.txt, IETF RFC 2396, August 1998.

[RFC2732]   R. Hinden, B. Carpenter, L. Masinter, *Format for Literal IPv6 Addresses in URL's*, http://www.ietf.org/rfc/rfc2732.txt, IETF RFC 2732, December 1999.

[XLink]   S. Bradner, *KXML Linking Language (XLink)*, http://www.w3.org/TR/2001/REC-xlink-20010627, W3C Recommendation, June 2001.

[XMLSchema]   P. V. Biron, A. Malhotra, *XML Schema Part 2: Datatypes*, http://www.w3.org/TR/2001/REC-xmlschema-2-20010502, W3C Recommendation, May 2001.

Appendix A. Revision History

[This appendix is optional, but helpful.]

Rev	Date	By Whom	What
0.1	June 29, 2004	Claus von Riegen	Initial Draft
0.2	September 20, 2004	Andrew Hately	First edit pass.
1.0	September 21, 2004	TC	First Version

Appendix B. Notices

OASIS takes no position regarding the validity or scope of any intellectual property 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; neither does it represent that it has made any effort to identify any such rights. Information on OASIS's procedures with respect to rights in OASIS specifications can be found at the OASIS website. Copies of claims of rights made available for publication 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 implementors or users of this specification, can be obtained from the OASIS Executive Director.

OASIS invites any interested party to bring to its attention any copyrights, patents or patent applications, or other proprietary rights which may cover technology that may be required to implement this specification. Please address the information to the OASIS Executive Director.