



Is your XML processor psychic?

Shedding a Little Light on XML

I got an e-mail the other day from Miss Cleo. It was quite exciting. She said she had a dream about me and that I needed to call her (or one of her psychic friends) to find out some important information about my future. Wow! I thought. So I ran to the phone.

We chatted for some time and I found out my girlfriend was cheating on me. How did she know? Those psychics amaze me. While I was moving out I thought how much this relates to XML namespaces.

XML, as you know, allows us to work with XML documents that have many different vocabularies. Elements and attributes may have identical names from different document models in a single XML document. Unless your XML processor has psychic abilities, it will need some additional information to distinguish between these elements and attributes and to understand their meaning correctly. This is where namespaces come into play.

Although the namespace recommendation has made this task easier for us, there are still a few areas of confusion. This month's column covers – and sheds a little light on – some of the very common questions on namespaces.

Q: Why do namespaces use a URL naming convention if they don't reference a Web site?

A: The Web has enjoyed the success of utilizing a naming convention that allows for unique file names regardless of the enormous number of files on the Internet. It's amazing if you think about it, yet still no problem with name resolution. If this concept had been considered before the Internet blossomed, the question of how to make certain that file names are unique would have been a huge challenge. Today it's simply a matter of creating a domain for your "corner of the Web." Most of us take this for granted.

A URL-formatted string is the most common form of a URI used for namespaces. URI strings are used in the W3C Namespace recommendation ("Namespaces in XML 1.0" at www.w3.org/TR/

[1999/REC-xml-names-19990114](http://www.w3.org/TR/1999/REC-xml-names-19990114)) to guarantee uniqueness in element type and attribute names when mixing vocabularies.

Any URI can be used for a namespace URI string, and the URL form is one of those allowed strings. The protocol identifier at the start of the URI dictates how the rest of the URI string is interpreted. The protocol identifier "http:" indicates that the next portion of the string is a domain name, and the "ownership" of domain names is governed. The remainder of the URI string can be any valid sequence of characters chosen by the owner of the domain. Provided that no one uses someone else's domain without permission, uniqueness is guaranteed between users.

What if a single document contains two elements with the same name? It's possible to use two different vocabularies within the same document, with element-type names used for different purposes. How can we distinguish them? This is precisely the problem that namespaces were created to solve.

This is also where the URL naming convention comes into play. The philosophy behind this is that URL naming convention has allowed us to define unique Internet filenames; why not use them to enforce unique element names in our XML documents? Similar problem? Yes. Similar solution? Yes. Similar success? Perhaps.

Well, this demonstrates a double-edged sword. Essentially, we've been conditioned to expect a resource at the end of the URL and first-time users find it a source of great confusion when looking for that "XML Library" or some other resource at the location specified by the URL. Take, for example, a namespace defined in an XSLT document:

```
<xsl:stylesheet xmlns:xsl="http://www.w3.org/1999/XSL/Transform" version="1.0">
```

On the surface it may appear that this namespace reference is pointing to the foregoing URL for reference to a possible library, a class, or even a definition of the current XSL spec. This is a common cause of confusion because we're conditioned to expect to retrieve *something* when we see a URL.

Admittedly, I made this invalid assumption myself at first. Yes, it's true, and I'm not afraid to admit it! Following is an excerpt of a conversation I had when I first started out:

Me: Are you *sure* it doesn't point to any type of library or anything required to make my XML work?

Expert: Yes.

Me: You're not lying to me, are you?

Expert: No.

Me: Why would it use a URL if it doesn't point to anything?

Expert: URLs are great at ensuring globally unique names.

Me: Will it fail if I'm not connected to the Internet at the time? [I thought I had him on this one.]

Expert: No, your document doesn't actually look at the resource. It just needs to follow that naming convention.

Me: Can I have a cup of coffee now?

Expert: Yes, I think you'll need it.

Bottom line: In Namespaces 1.0 the specification of URI syntax allows for globally unique names.

Q: What's the difference between a URL, a URN, and a URI?

A: Well, we're all too familiar with URLs, but a few other flavors of the UR* family have cropped up recently. Although they share a related goal of identifying resources, each has some definite distinctions. Here are some TLAs (three-letter acronyms):

- **URI:** Uniform Resource Identifier
- **URL:** Uniform Resource Locator
- **URN:** Uniform Resource Name

AUTHOR BIO

David Silverlight is chief XML evangelist for Infoteria Corporation (www.infoteria.com), an XML software development company based in Tokyo and Beverly, MA. He also maintains www.XMLPitstop.com, a resource for XML examples and everything else XML.

A URI is the broadest form of the term to describe a resource. In practice, the terms URI and URL are interchangeable; the primary distinction is that a URI merely identifies a resource whereas a URL identifies it and describes where and how to retrieve it. A URI, in fact, is a more generalized version of a URL, since it can refer to both a specific resource and a relative resource. You'll also find that certain characters permitted in URIs aren't permitted in URLs.

Let's start with Figure 1, a diagram that demonstrates how they're related.

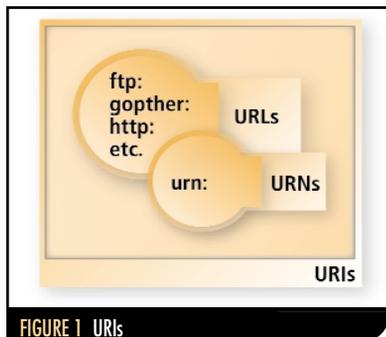


FIGURE 1 URIs

From Figure 1 you can see that URLs are a subset of URIs. Referring to a resource as a URL is a bit of an antiquated reference, yet it still gets the point across for users who are new to the UR* family. Actually, you might be revealing that you're technically over the hill if you refer to a resource as a URL in a technical document. URLs are most commonly used in conjunction with a subset of technologies such as http, ftp, and gopher. This first part of a URL, referred to as the *scheme*, tells an application what protocol to use in order to retrieve the resource.

A URN is a globally unique, location-independent resource name. The location independent nature is especially important here. If you change the name of the server where your URN resources are located, you won't break all of the links associated with it the way you would with a URI or URL. Much the same way that an ISBN number is associated with identifying a book, if you change the location of a book, the ISBN number won't change and you can still find it.

Microsoft's Hailstorm and .NET suite of products make great use of URNs and I suspect we'll see much more of them in the future.

Q: Why should I use namespaces in my XML?

A: That's a good question and it has many answers. The "X" in XML stands for *extensible*. Without namespaces, extending an XML vocabu-

lary is a difficult task. It's easy to add your own custom documentation tags to XSLT because your namespace won't collide with the XSLT namespace and confuse whatever XSLT engine you're using.

Another good reason is that one of the best things about XML is the ease with which data can be shared between applications. Without namespaces, sharing data becomes next to impossible. Applications have to make assumptions about the nature of the data they receive rather than rely on the namespace to identify data they can handle and data they can't or don't need to process.

[(This could be a whole article in and of itself.)]

Q: Where are namespaces headed?

A: Although the URI string in Namespaces 1.0 is *not* used for resource discovery, considerations are underway to publish a future version of Namespaces in which the URI string can be used for such a purpose. The Resource Directory Description Language (RDDL – www.rddl.org/) provides a text description of classes of resources, and of individual resources within those classes. RDDL is suitable for describing schemas, stylesheets, executable code, or any other resource associated with a vocabulary associated with a namespace URI. It has been proposed to embed RDDL descriptions in an XHTML document dereferenced from the namespace URI providing both human-readable and machine-readable information about a class of documents that are instances of namespace-identified vocabularies.

Elements of Design

While we're on the topic of namespaces, I recently discovered a great utility for maintaining, editing, and tracking XML namespaces. This tool, Namespaces Navigator, gives you the ability to add, edit, and track all the namespaces in your development environment.

The editing environment is very accommodating, as it allows you to view and/or update any section of your namespace. Pretty cool stuff. I'm glad somebody thought of this. What's more, it's built on a framework that allows for building entire apps on XML, XSLT, and HTML. The framework is another intriguing aspect of this tool and can be an educational process all on its own. Namespaces Navigator is actually an application of this framework. You can check it out at www.topxml.com/itse. ☺

DSILVERLIGHT @ INFOTERIA.COM

SUBSCRIBE AND SAVE

XML JOURNAL

Offer subject to change without notice

ANNUAL NEWSSTAND RATE

~~\$83.88~~

YOU PAY

\$77.99

YOU SAVE

\$5.89 Off the Newsstand Rate

DON'T MISS AN ISSUE!

Receive 12 issues of *XML-Journal* for only \$77.99! That's a savings of \$5.89 off the annual newsstand rate.

Sign up online at www.sys-con.com or call 1-800-513-7111 and subscribe today!

Here's what you'll find in every issue of XML-JOURNAL:

- Exclusive feature articles
- Interviews with the hottest names in XML
- Latest XML product reviews
- Industry watch

