2.4 Document Subsets

Some applications require the ability to create a physical representation for an XML document subset (other than the one generated by default, which can be a proper subset of the document if the comments are omitted). Implementations of XML canonicalization that are based on XPath can provide this functionality with little additional overhead by accepting a node-set as input rather than an octet stream. The processing of an element node ***E*** MUST be modified slightly when an XPath node-set is given as input and the element's parent is omitted from the node-set. This is necessary because omitted nodes SHALL not break the inheritance rules of inheritable attributes [C14N-Issues] defined in the xml namespace.

…, 11/5/07 2:52 PM
Deleted: some of the

…, 11/5/07 2:53 PM
Deleted: ancestors are

[Definition:] **Simple inheritable attributes** are attributes that have a value that requires at most a simple redeclaration. This redeclaration is done by supplying a new value in the child axis. The redeclaration of a simple inheritable attribute ***A*** contained in one of ***E***'s ancestors is done by supplying a value to an attribute ***Ae*** inside ***E*** with the same name. Simple inheritable attributes are `xml:lang` and `xml:space`.

The method for processing the attribute axis of an element ***E*** in the node-set is hence enhanced. All element nodes along ***E***'s ancestor axis are examined for the nearest occurrences of simple inheritable attributes in the xml namespace, such as `xml:lang` and `xml:space` (whether or not they are in the node-set). From this list of attributes, any simple inheritable attributes that are already in ***E***'s attribute axis (whether or not they are in the node-set) are removed. Then,

lexicographically merge this attribute list with the nodes of ***E***'s attribute axis that are in the node-set. The result of visiting the attribute axis is computed by processing the attribute nodes in this merged attribute list.

The `xml:id` attribute is not a simple inheritable attribute and no processing of these attributes is performed.

The `xml:base` attribute is not a simple inheritable attribute and requires special processing beyond a simple redeclaration. Hence the processing of ***E***'s attribute axis needs to be enhanced further. A "join-URI-References" function is used for `xml:base` fix up. It incorporates xml:base attribute values from omitted xml:base attributes and updates the xml:base attribute value of the element being fixed up, as follows.

An `xml:base` fixup is performed on an element E as follows. Let ***E*** be an element in the node set whose ancestor axis contains successive elements ***En***...***E1*** (in reverse document order) that are omitted and ***E***=***En+1*** is included. (It is important to note that ***En..E1*** is for contiguously omitted elements, for example only `e2` in the example in section 3.8.) The fix-up is only performed if at least one of ***E1*** ... ***En*** had an `xml:base` attribute. In that case let ***X1*** ... ***Xm*** be the values of the `xml:base` attributes on ***E1*** ... ***En+1*** (in document order, from outermost to innermost, ***m*** <= ***n+1***). The sequence of values is reduced in reverse document order to a single value by first combining ***Xm*** with ***Xm-1***, then the result with ***Xm-2***, and so on by calling the "join-URI-References" function until the new value for ***E***'s `xml:base` attribute remains. The result may also be null or empty (`xml:base=""`) in which case `xml:base` MUST NOT be rendered.

..., 11/5/07 5:23 PM
Deleted:

..., 11/5/07 5:36 PM
Deleted: ,

..., 11/5/07 5:36 PM
Deleted: which

..., 11/5/07 5:42 PM
Deleted: takes any URI (Base) from an ancestor and joins a relative URI of ***E*** (R) (in most cases after the last slash) of the former and then normalizes the result. We describe here a simple method for providing this functionality similar to that found in sections 5.2.1, 5.2.2. and 5.2.4. of RFC 3986 with the following modifications: <#>Perform RFC 3986 section 5.2.1. "Pre-parse the Base URI" modified as follows. <#>The scheme component is not required in the ... [1]

..., 11/5/07 5:43 PM
Formatted: Font:Courier, 12 pt

..., 11/5/07 5:35 PM
Deleted: This function may also be called with ... [2]

..., 11/5/07 5:21 PM
Formatted: Font:Bold, Italic

..., 11/5/07 5:20 PM
Formatted: Font:Courier

..., 11/5/07 5:39 PM
Deleted: n

..., 11/5/07 5:39 PM
Deleted: has

..., 11/5/07 5:23 PM
Deleted:

..., 11/5/07 5:40 PM
Deleted: described previously

Note that this xml:base fixup is only performed if an element with an xml:base attribute is removed. Specifically, it is not performed if the element is present but the attribute is removed.

The join-URI-References function takes an xml:base attribute value from an omitted element and combines it with other contiguously omitted values to create a value for an updated xml:base attribute. A simple method for doing this is similar to that found in sections 5.2.1, 5.2.2 and 5.2.4 of RFC 3986 with the following modifications:

- Perform RFC 3986 section 5.2.1. " Pre-parse the Base URI" modified as follows.
  - The scheme component is not required in the base URI (Base). (i.e. Base.scheme may be null)
  - Replace a trailing ".." segment with "../" segment before processing.

- 5.2.4. "Remove Dot Segments" is modified as follows:
  - Keep leading "../" segments
  - Replace multiple consecutive “/” characters with a single “/” character.
  - Append a “/” character to a trailing “..” segment

- The "Remove Dot Segments" algorithm is modified to ensure that a combination of two xml:base attribute values that include relative path components (i.e., path components that do not begin with a '/' character) results in an attribute value that is a relative path component.

- Perform RFC 3986 section 5.2.2. "Transform References" modified as follows to ignore the fragment

…, 11/5/07 5:42 PM
Formatted: Bullets and Numbering

11/6/07 10:54 AM
Formatted: Bulleted + Level: 1 + Aligned at: 0.25" + Tab after: 0.5" + Indent at: 0.5"

11/6/07 10:50 AM
Formatted: Bullets and Numbering

11/6/07 10:54 AM
Formatted: Font:16 pt

11/6/07 10:54 AM
Formatted: Bulleted + Level: 2 + Aligned at: 0.75" + Tab after: 1" + Indent at: 1"

11/6/07 10:54 AM
Formatted: Font:16 pt

11/6/07 10:54 AM
Formatted: Font:16 pt

11/6/07 10:54 AM
Formatted: Bullets and Numbering

11/19/07 6:12 PM
Formatted: Bulleted + Level: 1 + Aligned at: 0.25" + Tab after: 0.5" + Indent at: 0.5", Tabs: 0.65", Left + Not at 0.15"

11/6/07 11:06 AM
Formatted: Bullets and Numbering

part of R

- After parsing R set R.fragment = null

Then, lexicographically merge this fixed up attribute with the nodes of ***E***'s attribute axis that are in the node-set. The result of visiting the attribute axis is computed by processing the attribute nodes in this merged attribute list.

Attributes in the XML namespace other than `xml:base`, `xml:id`, `xml:lang`, and `xml:space` MUST be processed as ordinary attributes.

The following examples illustrate the modification of the "Remove Dot Segments" algorithm:

- "abc/" and "../" should result in ""
- "../" and "../" are combined as "../../" and the result is "../../"
- ".." and ".." are combined as "../../" and the result is "../../"

To illustrate the last example, when the elements b and c are removed from the following sample XML document, the correct result for the xml:base attribute on element d would be "../../x":

```
<a xml:base="foo/bar">
  <b xml:base="..">
    <c xml:base="..">
      <d xml:base="x">
      </d>
    </c>
```

11/12/07 8:32 AM
Formatted: Indent: Left: 0.75"

11/6/07 11:07 AM
Formatted: Space After: 0 pt, Tabs: 0.15", Left

11/19/07 6:20 PM
Formatted: Font:16 pt

11/19/07 6:20 PM
Formatted ... [3]

11/19/07 6:20 PM
Formatted: Font color: Auto

11/19/07 6:17 PM
Formatted: Font:16 pt

11/19/07 6:14 PM
Formatted: Bullets and Numbering

11/19/07 6:17 PM
Formatted: Font color: Auto

11/19/07 6:17 PM
Formatted: Font:16 pt

11/19/07 6:18 PM
Formatted: Indent: Left: 0.25"

11/19/07 6:17 PM
Formatted: Font:16 pt

11/19/07 6:17 PM
Deleted:

3.8 Document Subsets and XML Attributes

Input Document	`<!DOCTYPE doc [`
	`<!ATTLIST e2 xml:space (default|prese`
	`<!ATTLIST e3 id ID #IMPLIED>`
	`]>`
	`<doc xmlns="http://www.ietf.org" xmln`
	`xml:base="something/else">`
	`  <e1>`
	`  <e2 xmlns="" xml:id="`
	`  <e3 id="E3" xml:ba`
	`  </e2>`
	`  </e1>`
	`</doc>`

Document Subset Expression	`<!-- Evaluate with declaration xmlns:`
	
	`(//. | //@* | //namespace::*)`
	`[`
	`  self::ietf:e1 or (parent::ietf:e1`
	`  or`
	`  count(id("E3")|ancestor-or-self::n`
	`self::node())`
	`]`

Canonical Form	`<e1 xmlns="http://www.ietf.org" xmlns`
	`xml:base="something/else"><e3 xmlns="`
	`xml:space="preserve"></e3><`

..., 11/5/07 5:12 PM
Deleted: http://www.example.com/

..., 11/5/07 5:13 PM
Deleted: ../

..., 11/5/07 5:14 PM
Deleted: http://www.example.com/

..., 11/5/07 5:15 PM
Deleted: http://www.example.com

Demonstrates:

- `xml:id` not inherited.
- simple inheritable XML attribute inherited (`xml:space`)
- `xml:base` fixup performed

Appendix A

11/20/07 10:33 AM
Formatted: Font:20 pt, Bold

Use the material in the table in Appendix A following "Some Examples" as given in

11/20/07 10:33 AM
Formatted: Font:Italic

http://lists.w3.org/Archives/Public/public-xml-core-wg/2007Jun/att-0050/Apendix_20060625.html

Add the following text before table:

11/6/07 11:04 AM
Formatted: Font:Helvetica, 16 pt, Italic

The following informative table outlines example results of the modified Remove Dot Segments algorithm described in Section 2.4.