

# Re: Word 2003 XML

---

*Source:*

<http://www.tech-archive.net/Archive/Mac/microsoft.public.mac.office.word/2006-06/msg00311.html>

---

- *From:* "Peter Jamieson" <[pjj@xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx](mailto:pjj@xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx)>
  - *Date:* Tue, 20 Jun 2006 01:41:58 +0100
- 

My very limited understanding of XML extensibility is that it is made clear to the parser/application which elements are foreign by declaring namespaces in the root element.

I would see it slightly differently.

In essence the parser determines whether the XML provided to it, i.e.

- a. is "well-formed" (i.e. conforms to XML syntax rules, e.g. that elements begin with "<" and end with ">")
- b. is "valid" (i.e. conforms to any XML Schemas that the parser is obliged to honour)

The parser isn't expected to know what any of the XML "means." and it has no way of knowing what the application "understands" or what the application sees as "native" or "foreign."

It's up to the application what it makes of the XML coming from the parser.

I would agree that there is no general-purposes XML mechanism that tells an application what to do with XML it doesn't recognise, but that does not prevent an application from trying to do something sensible with that XML. It could do that by (for example)

- a. being provided with enough information in XML elements it /does/ understand to locate and invoke an application (cf. an olde worlde Microsoft Office applet) that knows what to do with the "alien" elements. That's more or less what Word does with embedded objects in .doc format documents, and there's no reason why the OLE/ActiveX mechanism could not be modelled using XML.
- b. trying to guess what application to use based on the XML content. Hazardous, no doubt!
- c. providing a standardised "I don't understand this content" response, e.g. by displaying a placeholder, or simply displaying "I couldn't understand the following XML" and quoting the XML, or some such.

It may also be worth pointing out that in the general case, a parser may not have much info. about the schemas associated with an XML document. The schemas, or rather the associated namespaces, may be named in the document (e.g. using attributes that define alias names), but the names are just

Re: Word 2003 XML

names, primarily intended to ensure uniqueness in the event of name clashes (which is what namespaces are all about). Even where the names look like URLs they do not necessarily point to any additional information about the schema. For example, one of the first snippets in a WordProcessingML document defines an alias "w" as follows:

```
<w:wordDocument  
xmlns:w="http://schemas.microsoft.com/office/word/2003/wordml
```

If you open the http address in your browser, there is nothing there – in fact I get a "page does not exist" error. There is another mechanism for providing information about the location of schema information to a parser, but as far as the XML standards are concerned, the mechanism merely provides a /hint/ – the parser can choose to go and find the schema, or ignore the hint. If an application wants a parser to honour the hint, it had better use a parser that can be relied on to do so.

Peter Jamieson

"Chris Ridd" <chrisridd@xxxxxxx> wrote in message  
news:4fnvfvlj5pdsU1@xxxxxxxxxxxxxxxxxxxx

On 2006-06-19 13:56:35 +0100, "John McGhie [MVP – Word and Word  
Macintosh]" <john@xxxxxxxxxxxx> said:

Yeah, I know DTDs are ancient technology. Their one saving grace is that they "always" work. I don't trust Schemas 100 per cent yet. They're lovely when they work, and Office 2007 supports (and uses) them.

I'm sure you're right. We can probably say "bah" and "humbug" at this point!

However, if I were working "cross platform" trying to make Office do the right thing with Open Doc, I would rather bet my life on doing it with a DTD and a FOSI than with a Scheme.

EMPTY properties, see <http://www.w3.org/TR/rdf-syntax-grammar/>

I think that might be specific to RDF, and not XML in general.

My very limited understanding of XML extensibility is that it is made clear to the parser/application which elements are foreign by declaring namespaces in the root element. I don't think any general spec says what an application's meant to do when alien elements or attributes are discovered in a document.

Re: Word 2003 XML

If you have foreign stuff in a document, validation through a DTD becomes impossible. I've a feeling that you may be able to validate such a document using a schema?

I fully anticipate being corrected on some or all of the above!

Cheers,

Chris

.